

Les écarts à l'indépendance.
Techniques simples pour analyser des
données d'enquête.

Philippe Cibois
Professeur de sociologie à l'université
de Versailles - St-Quentin en Yvelines

VISIT...

LANZAROTE
Caliente.COM

En hommage à G. Th. Guilbaud

Introduction

Vous étudiez la sociologie (en formation initiale dans un département de sociologie, d'AES ou de Sciences politiques ; en formation continue ou par intérêt personnel) et vous êtes confrontés dans vos lectures (ou dans vos cours) à l'utilisation de techniques statistiques appliquées à des données d'enquête. Ce livre voudrait vous aider à comprendre en quoi consistent, dans l'esprit de cette collection, certaines de ces *statistiques appliquées à la sociologie* (pour reprendre l'intitulé du programme du Deug de sociologie). Je voudrais dans ce texte introductif vous expliquer en quoi consiste la spécificité de cette présentation, ce que l'on trouvera dans ce livre, comment l'utiliser et ce qu'on n'y trouvera pas.

Une présentation tournée vers l'utilisateur

Les étudiants en sociologie n'aiment pas en général les mathématiques pour une raison simple, c'est que, comme le montre la sociologie de l'éducation, ils ont été contraints de s'orienter pendant leur secondaire soit vers les lettres, soit vers les sciences économiques et sociales, soit dans d'autres directions du fait de leur faible réussite en mathématiques. Pour arriver à faire ce tri social, l'enseignement des mathématiques est dans le secondaire d'un niveau d'abstraction fort, voulu et efficace : il arrive à persuader beaucoup de gens qu'ils sont *nuls en math*, que *ce n'est pas leur truc* ou autres justifications après coup qui tentent de rendre compte de ce qui est ressenti souvent comme une humiliation ou un échec.

Il s'agit là d'une spécificité française d'ailleurs relativement récente et que je ne souhaite pas éternelle : dans beaucoup de pays,

l'enseignement des mathématiques est destiné simplement à apprendre les mathématiques et non à sélectionner. De ce fait les étudiants du monde anglo-saxon qui arrivent en sociologie ont beaucoup moins de difficultés à apprendre les techniques quantitatives que leurs homologues français qui sont souvent bloqués devant toute présentation qui leur rappelle de trop mauvais souvenirs.

Si vous vous reconnaissez dans ce tableau, vous n'avez pas à culpabiliser mais simplement à accepter de voir les choses différemment. Car pour s'adresser à vous, il faut renouveler la manière de faire et tout simplement prendre celle en vigueur dans les pays où les mathématiques servent simplement à apprendre les mathématiques. Quand on prend l'usage anglo-saxon d'exposition des statistiques, on fait en sorte de s'adresser à des utilisateurs de statistiques en vue d'en faire des utilisateurs avertis, non pour les former subrepticement aux mathématiques.

La technique pédagogique est simple, elle se trouve par exemple dans les nombreux ouvrages de la collection *Quantitative Applications in the Social Sciences* de l'éditeur Sage où en général pour exposer une technique statistique, on part d'un exemple numérique, on montre ce qu'apporte la technique aux sciences sociales et on donne des compléments pour formaliser le cas échéant. Quand on présente un exemple numérique, l'utilisateur n'est pas à ce point borné qu'il ne comprenne pas que le cas présenté soit général et que la technique soit adaptée à d'autres cas concrets.

Dans ce qui va suivre, l'exposé ne cherchera pas à valider formellement les techniques statistiques : vérifier que tout est en ordre de marche, sans vice caché, est le travail du mathématicien et du statisticien. Ce sont eux qui en tant que professionnels contrôlés par leurs pairs sont à même de nous donner les instruments que nous utiliserons. Notons cependant que parmi les grands noms de la statistique, Pearson, Fisher, Spearman, tous ont été tournés vers les

applications et que le dernier était d'abord un psychologue, ce qui ne l'a pas empêché d'être le créateur d'une technique qui joue aujourd'hui un grand rôle dans le dépouillement d'enquête (l'analyse factorielle). La validation d'une technique est une chose, son utilisation correcte une autre : quand nous avons appris à faire une division, nous n'avons pas appris ce qui rend valide cette opération mais son fonctionnement, son champ d'application et ses pièges.

La présentation qui va suivre est tournée vers l'utilisateur en ce sens que les techniques statistiques appliquées à la sociologie ne sont bien comprises que si on en voit l'intérêt, c'est à dire si, à partir d'exemples réels, on montre en quoi leur utilisation aide à la compréhension sociologique. De ce fait, il vaut mieux être sociologue pour enseigner la statistique aux sociologues.

L'analyse des données d'enquête

Cet ouvrage se veut d'introduction, il présente des *techniques simples*, ce qui qualifie un niveau de compréhension, et d'introduction à *l'analyse des données d'enquête*, ce qui désigne un type de données. En un mot il s'agit de l'analyse des tableaux croisés obtenus à partir de données d'enquête.

La procédure d'enquête commence à faire partie du bagage commun des savoirs. Il s'agit d'interroger un nombre limité de personnes : 25, 50, 100, 500, 1000 ou plus, et de poser aux interrogés des questions sur leurs pratiques, leurs opinions, leur situation. Les renseignements obtenus sont sous forme de *modalités* de réponses : par exemple la *question* "sexe" est une question à deux *modalités* de réponse "masculin", "féminin". Souvent les réponses recueillies sous forme purement numérique (âge, revenu, temps passé à une activité) sont ramenées également à des modalités, par exemple sous forme de tranches d'âges : "18-24ans", "25-30ans", etc. Dans ce qui suit on supposera que les données d'enquêtes sont toutes ramenées à des modalités de réponse à des questions. On laisse à un autre ouvrage le

soin de traiter des statistiques appliquées aux données numériques comme telles (ouvrage qui constituerait par exemple le deuxième semestre d'un cours de première année de Deug de sociologie, le présent ouvrage en constituant le premier semestre).

Pour que tout soit bien clair, précisons le fonctionnement de la procédure d'enquête : par exemple demandons aux 225 étudiants d'un amphi de Deug 1 de sociologie de répondre rapidement sur une feuille aux deux questions suivantes : "fumez-vous au moins une cigarette par jour ?" (deux modalités : "oui"/"non") et "quel est votre sexe ?" ("masculin"/"féminin"). A la fin du cours, quatre tas de réponses pouvaient être faits, d'épaisseurs inégales : un premier tas de 29 feuilles qui représente les individus qui étaient à la fois masculin et fumeurs (ayant répondu "oui" à la question sur le fait de fumer), puis viennent les 17 masculin et non-fumeurs, les 70 féminin ne fumant pas et les 109 féminin et ayant répondu oui à la question sur le fait de fumer. Faire un tableau croisé consiste à répartir ces 4 nombres en un tableau à deux dimensions où par exemple les lignes signalent les modalités de sexe et les colonnes les modalités de réponse au fait de fumer. En totalisant les lignes et les colonnes on a le tableau croisé suivant qui est classique et sur lequel nous reviendrons plus loin (Enquête notée FUM97).

	Oui fume	Non ne fume pas	<i>Total</i>
Sexe masculin	29	17	46
Sexe féminin	70	109	179
<i>Total</i>	99	126	225

Les *marges* du tableau sont les *nombres en italique* obtenus par totalisation : la marge colonne nous indique que sur 225 étudiants (total général en gras), 46 sont de sexe masculin et 179 de sexe féminin et la marge ligne nous indique déjà que les non-fumeurs sont plus nombreux que les fumeurs.

Le *tableau croisé* est l'outil de base du sociologue dès qu'il est confronté à des données d'enquête : certes le sociologue qui analyse ce type de données commence toujours par regarder chaque question individuellement par le biais de la répartition des effectifs de chaque modalité mais rapidement, on éprouve le besoin de croiser des questions. Les techniques pour aider à la lecture des tableaux croisés seront étudiées dans le premier chapitre.

Le deuxième chapitre tentera de répondre à la question suivante : les 225 de cet amphi de Deug sont-ils suffisamment nombreux pour trancher la question de savoir s'il y a une liaison entre le sexe et le fait de fumer. Si on voit une liaison dans le cas présent, est-elle généralisable à d'autres amphis comparables ?

Les deux chapitres de cet ouvrage rendent compte des deux questions de base que se pose un sociologue face à un tableau croisé qui cherche à savoir : ce que dit le tableau ; si les résultats aperçus sont fiables. D'autres techniques utilisées par le sociologue pour dépouiller des enquêtes sont à envisager : elles permettent de prendre ensemble en considération non pas deux questions de l'enquête mais davantage. Ce sont les techniques d'analyse factorielle et de régression sur données d'enquête : elles dépassent le cadre du présent ouvrage.

Comment utiliser ce livre

La technique d'exposition partant toujours d'un exemple, il faut refaire les calculs mais il faut surtout appliquer les méthodes sur d'autres données. Le présent exposé ne trouve pleinement sa valeur qu'associé à des travaux pratiques sur des données d'enquêtes. Il est en général assez facile de s'en procurer car les enquêtes faites à des fins pédagogiques et de recherche sont nombreuses. Pour le traitement, le recours à l'informatique semble aujourd'hui indispensable. Des tableaux croisés peuvent être obtenus avec tous les logiciels qui traitent de données d'enquêtes. Certaines techniques plus spécifiques sont disponibles sur plusieurs de ces logiciels (Trideux, Modalisa).

Sources

La technique de lecture des tableaux croisés en utilisant les pourcentages en ligne n'est pas à ma connaissance attribuée à un auteur particulier : elle fait partie du savoir commun d'une discipline. J'ai proposé la technique de lecture rapide dans *L'analyse des données en sociologie*, paru en 1984 aux Presses Universitaires de France ; la représentation des écarts à l'indépendance a été présentée d'une manière formalisée dans "Méthodes post-factorielles pour le dépouillement d'enquête", *Bulletin de Méthodologie sociologique*, n°1, octobre 1983 et les indices de liaison d'une attraction dans "Le PEM, pourcentage de l'écart maximum : un indice de liaison entre modalités d'un tableau de contingence", *Bulletin de méthodologie sociologique*, n°40, septembre 1993.

Quant au Khi-deux, il est centenaire puisqu'il a été proposé par Karl Pearson en 1900. Dans le titre de la publication d'origine¹, il propose de trouver une manière de faire "raisonnable" pour tester si une liaison observée est due ou non au hasard. Il n'est pas sûr que depuis cette date, la littérature qui traite du Khi-deux soit restée "raisonnable", surtout dans les manuels qui l'enseignent. On essaiera ici, dans la ligne du créateur et de certains de ses successeurs de proposer une pratique raisonnable de cet indicateur².

¹ "On the Criterion that a given System of Deviations from the Probable in the case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling", *The London, Edinburgh and Dublin Philosophical magazine and Journal of Science*, 1900, vol. 50, p.157-175. Réédité d'une manière plus accessible dans *Kearl Pearson early Statistical Papers*, Cambridge University Press, 1956.

² Ce texte a été orthographié en suivant les recommandations du Conseil supérieur de la langue française approuvées à l'unanimité par l'Académie française le 3 mai 1990 et que celle-ci a souhaité, le 17 janvier 1991 soumettre "à l'épreuve

Chapitre 1

Attraction et indépendance

1 L'attraction

L'enquête 1989 sur les pratiques culturelles des français³ nous apprend que les 4997 enquêtés se répartissent de la manière suivante si l'on considère le sexe et le fait d'aller à la chasse⁴ :

Hommes allant à la chasse	172
Femmes " "	23
Hommes n'allant pas à la chasse	2232
Femmes " "	2570

Total	4997

De cette distribution d'effectifs nous pouvons d'abord tirer le pourcentage de ceux qui vont à la chasse : 172 hommes + 23 femmes soit 195 individus sur un total de 4997, soit donc une *proportion* de $195/4997 = 0,039$ et un *pourcentage* de 3,9%. Il s'agit donc d'une pratique assez rare.

Notons qu'une *proportion* est toujours comprise entre 0 et 1, un *pourcentage* entre 0 et 100 : on calcule en proportion parce que c'est

du temps". Ces modifications portent surtout sur des simplifications d'accent (absence d'accent circonflexe sur le "i" et sur le "u" sauf cas d'ambiguïté et régularisations : "évènement" comme "avènement").

³ *Nouvelle enquête sur les pratiques culturelles des français en 1989*, Paris, 1990, La Documentation française, désignée dans la suite PRAT89.

⁴ Aller à la chasse souvent ou de temps en temps

en général plus simple, mais on parle en pourcentage parce que depuis l'époque contemporaine, c'est le mode le plus habituel pour comparer des grandeurs. Pour passer d'une représentation à l'autre, on déplace la virgule de deux rangs. Il faut se souvenir que l'expression 3,9 *pour cent* est synonyme de $3,9/100$, c'est à dire de la proportion correspondante de 0,039.

A partir de cette distribution, nous pouvons également répondre à la question suivante : « qui va le plus à la chasse ? Les hommes ou les femmes ? ». La réponse semble évidente mais précisons la en disposant la distribution sous forme d'un *tableau croisé*⁵, où à chaque ligne correspond une catégorie de sexe et à chaque colonne le fait d'aller à la chasse ou non. Sur les marges du tableau sont disposées les sommes des lignes et des colonnes

	Chasse	Non chasse	Total
Hommes	172	2232	2404
Femmes	23	2570	2593

Total	195	4802	4997

Calculons pour chaque ligne le pourcentage de chaque sexe dans les deux catégories et le pourcentage moyen (déjà calculé)

	Chasse	Non chasse	Total
Hommes	7,2	92,8	100
Femmes	0,9	99,1	100

Total	3,9	96,1	100

⁵ appelé aussi *tableau de contingence*. Ce terme s'emploie surtout dans les milieux de la statistique (mathématiques ou économie). En sociologie c'est plutôt le terme de tableau croisé qui est employé.

La chasse est une pratique rare (environ 4%) mais très peu fréquente chez les femmes (moins de 1%) et proche de 7%⁶ chez les hommes. Le pourcentage moyen de 3,9% nous sert de repère et la comparaison se fait entre ce pourcentage et ceux de chacune des deux catégories de sexe : il est plus fort que la moyenne pour les hommes, plus faible pour les femmes. On compare toujours perpendiculairement au pourcentage en ligne.

Ce résultat n'est pas surprenant : chacun sait que la chasse est une activité socialement *marquée* comme masculine. Ce qu'il est intéressant de voir ici c'est que nous pouvons, par la comparaison des pourcentages, formaliser ce marquage social. C'est ce que nous appellerons désormais *l'attraction*. L'attraction entre deux réalités sociales est révélée par la différence au pourcentage moyen. Ici il y a attraction entre le fait d'aller à la chasse et d'être de sexe masculin.

En cherchant dans la même enquête des activités marquées cette fois comme féminines, on retrouve le même phénomène d'attraction avec le fait de faire soi-même des vêtements.

⁶ $172/2404 = 0,07154742097....$ et il faut choisir une précision. Ici, sur un total correspondant à 2404, 1% correspond à 24 personnes ce qui est un manque manifeste de précision, 0,1% à 2,4 personnes et 0,01% à 0,24 personne, ce qui devient incohérent car la précision minimum ne peut être que d'une personne et pas d'un quart de personne. On choisit donc la précision d'un chiffre après la virgule pour la précision des pourcentages. On conservera cette précision même si les effectifs de référence sont inférieurs afin de toujours distinguer typographiquement entre les *effectifs observés* qui sont des entiers et les *pourcentages* qui seront toujours décimaux. Quant à l'arrondi, il se fera toujours *au plus près*, non pour une précision illusoire, mais pour répartir les restes tantôt par excès et tantôt par défaut (on écrira 7,1 pour 7,10 à 7,14 et 7,2 pour 7,15 à 7,19). Le total de 100 est une référence qui doit être affichée comme telle, non la somme des arrondis qui peut très bien différer de 100.

	Fait des vêtements	N'en fait pas	Total
Hommes	33	2371	2404
%	1,4	98,6	100
Femmes	472	2121	2593
%	18,2	81,8	100

Total	505	4492	4997
%	10,1	89,9	100

En moyenne 10% de la population participe à la fabrication de ses vêtements mais la proportion⁷ en est plus forte pour les femmes (18,2%) que pour les hommes (1,4%). On peut ajouter cette information simplifiée sur le tableau en notant d'un signe *plus* les écarts positifs au pourcentage moyen et d'un signe *moins* les écarts négatifs. On dispose ainsi d'une *lecture rapide* du tableau

	Fait des vêtements	N'en fait pas
Hommes	-	+
Femmes	+	-

Il y a attraction entre « faire ses vêtements » et sexe féminin et opposition entre « faire ses vêtements » et sexe masculin (on a inversion des signes pour la deuxième colonne). On notera qu'une opposition n'est pas une absence : des hommes font des vêtements, mais moins que la moyenne.

Que se passe-t-il si l'on considère les pourcentages en colonnes : cela change-t-il les résultats et leur interprétation ?

Fait des vêtements	N'en fait pas	Total
--------------------	---------------	-------

⁷ on suit le langage courant qui parle spontanément de *proportion*, mais on visualise cette proportion par son *pourcentage* correspondant. Ce double langage sera souvent employé dans la suite.

	%	%	%
Hommes	33	2371	2404
	6,5	52,8	48,1
Femmes	472	2121	2593
	93,5	47,2	51,9

Total	505	4492	4997
	100	100	100

En termes d'attractions ou d'oppositions les résultats sont les mêmes mais les références de comparaisons sont différentes : la moyenne est ici la répartition des sexes, un peu en faveur des femmes, ce qui est normal du fait de la surmortalité masculine pour une enquête qui prend en compte toute la population à partir de 18 ans. L'attraction « femmes - vêtements » est repérable car 93,5% (proportion des femmes dans ceux qui font leurs vêtements) est plus fort que 51,9% (proportion moyenne des femmes). Tous les signes des écarts au pourcentage moyen sont les mêmes bien que tous les pourcentages soient différents.

L'inconvénient de cette deuxième présentation du tableau est qu'elle masque le fait que la fabrication de vêtements est un phénomène plutôt rare, quelque soit le sexe, ce qui était bien mis en relief dans la première présentation puisque c'était la base de la comparaison. Par contre, ce qui sert de référence dans cette deuxième présentation, c'est la comparaison des sexes qui n'est pas le phénomène que l'on souhaite étudier car c'est déjà quelque chose de connu. On trouve là la justification de la pratique habituelle des sociologues de mettre par convention en ligne la variable préalablement connue (variable dite aussi « explicative ») et de mettre en colonne la variable nouvelle, celle dont on veut rendre compte

(variable dite « à expliquer »), et de toujours calculer les pourcentages *en ligne*.

Que l'on travaille avec des pourcentages en ligne ou en colonne, et c'est un résultat général, le signe des écarts au pourcentage moyen est le même : il manifeste pour un signe *plus* l'attraction entre la modalité de ligne et celle de colonne (ou son opposition pour un signe *moins*). C'est un *invariant* du tableau, il ne dépend pas de ce qui est mis en ligne ou en colonne et comme tel il doit être privilégié.

L'attraction, quand elle est suffisamment forte est déjà repérée par le sens commun car certaines activités sont marquées socialement et perçues comme telles : l'examen du tableau croisé et des pourcentages en ligne permet de formaliser la présence de cette attraction. Avant d'en étudier la force il faut envisager le cas de l'absence d'attraction.

2 L'indépendance

Toujours de la même enquête on peut tirer le tableau croisé entre le sexe et le fait de s'occuper d'un jardin : on donne les effectifs et les pourcentages en ligne.

	S'occupe d'un jardin		Total
	Oui	Non	
Hommes	965	1439	2404
%	40,1	59,9	100
Femmes	1052	1541	2593
%	40,6	59,4	100
Total	2017	2980	4997
%	40,4	59,6	100

En moyenne environ 40% des gens s'occupent d'un jardin et cette proportion moyenne est celle des deux sexes : cette absence de différence signifie qu'il n'y a pas d'attraction ni d'opposition entre la pratique étudiée et le sexe : il y a *indépendance entre les deux variables*. On retrouve le même résultat en comparant les pourcentages de la deuxième colonne.

Cependant, si l'on regarde les choses de près, on doit dire que les hommes s'occupent moins d'un jardin que les femmes puisque ils s'en occupent pour 40,1% d'entre eux contre 40,6% chez les femmes. Mais cet écart est-il notable ou négligeable ? Pour répondre à cette question, mettons nous dans le cas de l'indépendance stricte, de l'égalité des pourcentages de lignes au pourcentage moyen. Si la situation moyenne de 40,4% de pratique du jardin s'appliquait strictement aux 2404 hommes, combien serait-il exactement à s'occuper de leur jardin ?

40,4% correspond à une proportion plus exacte⁸ de $2017 / 4997 = 0,40364$ qui appliquée par multiplication au 2404 hommes donne un effectif de 970,4. Cet effectif, du simple fait qu'il ne soit pas un entier est évidemment hypothétique et c'est pourquoi on l'appelle *effectif théorique* correspondant à la situation hypothétique d'indépendance⁹.

L'écart entre l'effectif observé et l'effectif théorique est de $965 - 970,4 = -5,4$ individus ce qui est faible et peut être considéré comme négligeable car une différence d'environ 5 individus par rapport à une population observée de 965 est peu de choses. Le calcul est analogue pour les autres cases où l'on voit que les différences positives ou négatives par rapport à l'effectif théorique correspondant à l'indépendance sont de 5,4 individus. Au total c'est donc environ une

⁸ pour ne pas avoir une dérive d'arrondis dans les calculs, on prend ici pour les opérations davantage de décimales

⁹ ou, pour reprendre une terminologie courante, l'effectif théorique sous l'hypothèse d'indépendance

vingtaine d'individus qui est en écart à l'indépendance (soit en positif, soit en négatif, ce qui par rapport à une population totale d'environ 5000 individus représente $20 / 5000 = 0,004$ soit 0,4%). On peut donc conclure qu'il n'y a pratiquement pas d'écarts entre ce qui est observé et ce qu'il y aurait s'il y avait indépendance entre les deux variables. Faire son jardin n'est pas une opération marquée par le sexe, c'est une opération aussi bien faite par des hommes que par des femmes.

On remarquera que :

1) la proportion de 0,40364 est le rapport du nombre de personnes qui s'occupent d'un jardin (égal à 2017) par rapport à l'ensemble de la population de 4997. C'est ce rapport de $2017/4997$ que l'on multiplie par le total du nombre d'hommes de 2404 : on peut résumer l'opération en disant que le produit $2017 / 4997 \times 2404$ ou encore $2017 \times 2404 / 4997 = 970,4$ est le produit des effectifs des marges (s'occuper d'un jardin et sexe masculin) divisé par le total. D'une manière sténographique on dit que *l'effectif théorique se calcule en faisant le produit des marges divisé par le total*¹⁰.

2) ce calcul est symétrique. Si l'on travaille en colonnes, la proportion moyenne est de $2404 / 4997$ que l'on applique au 2017 qui s'occupent de leur jardin et l'on retrouve le même résultat de 970,4 comme effectif théorique. Effectif théorique et effectif en écart à l'indépendance sont également des *invariants* du tableau, c'est à dire qu'ils ne dépendent pas de sa disposition concernant ce qui est mis en ligne ou en colonne.

3) dans ce tableau, une autre manière de repérer que l'écart à l'indépendance est faible est de voir ce qui se passerait si les données

¹⁰ Formule à retenir : théorique = produit des marges / total ; ou produit des mages sur total, ou encore produit des marges par total (sous-entendu divisé par)

étaient très peu différentes, par exemple si 8 hommes de plus étaient dans le camp de ceux qui s'occupent de leur jardin et 8 femmes dans celui de ceux qui ne s'en occupent pas¹¹. Les marges du tableau seraient ainsi identiques et donc les effectifs théoriques seraient les mêmes. Dans ce cas, pour la case masculin-jardin, l'observé (imaginaire) est de $965 + 8 = 973$, et le pourcentage des jardiniers chez les hommes est désormais de $973 / 2404$ soit 40,5%, ce qui est supérieur au pourcentage moyen. L'attraction se fait maintenant entre sexe masculin et le fait de jardiner. Le même phénomène est repéré en comparant l'effectif observé (imaginaire) de 973 avec l'effectif théorique. Cette fois la différence effectif observé moins effectif théorique est de $973 - 970,4 = 2,6$ individus.

Cet écart positif est le reflet de cette attraction (imaginaire). En conclusion si donc le changement de quelques individus a entraîné l'inversion des signes des écarts, c'est que ces signes sont ici peu fiables du fait qu'on est très proche de l'indépendance.

3 L'écart à l'indépendance

Puisque l'effectif théorique est un invariant ne dépendant pas de l'orientation ligne/colonne, il est possible de calculer pour chaque case (des données réelles et non imaginaires) l'effectif correspondant à l'indépendance (produit des marges par le total), puis par différence avec l'observé¹², l'écart à l'indépendance. On retrouvera pour la

¹¹ Attention : il s'agit d'une expérience fictive pour voir ce qui se passerait si on était dans cette situation. La situation observée est ce qu'elle est et on n'a jamais le droit de la modifier : une expérience "de pensée" doit rester virtuelle. Les données sont intouchables.

¹² abréviation classique d'effectif observé ; on dit aussi l'observation.

première case ce que nous avons déjà donné : observé = 965 ;
théorique¹³ = 970,4, écart¹⁴ = -5,4

	S'occupe d'un jardin		
	Oui	Non	Total
Homme : observé	965	1439	2404
théorique	970,4	1433,6	
écart	-5,4	+5,4	
Femme : observé	1052	1541	2593
théorique	1046,6	1546,4	
écart	+5,4	-5,4	

Total	2017	2980	4997

Les quatre écarts à l'indépendance sont identiques en valeur absolu : ce phénomène est lié au format du tableau, le fait qu'il soit à 2 lignes et 2 colonnes. Quand l'une de ces dimensions devient supérieure à 2, le phénomène d'égalité en valeur absolue disparaît pour cette dimension.

Prenons dans la même enquête un tableau où les différences soient plus marquées : il suffit pour cela de ne pas prendre la pratique du jardinage en général, mais le type de jardinage qui dans l'enquête était distingué en deux catégories : jardin potager et jardin d'agrément. Soit le même croisement pour la pratique du seul jardin d'agrément (fleurs, pelouse) :

¹³ abréviation de effectif théorique sous l'hypothèse d'indépendance

¹⁴ abréviation d'effectif d'écart entre l'observé et le théorique

S'occupe d'un jardin d'agrément			
	Oui	Non	Total
Hommes	745	1659	2404
%	31,0	69,0	100
Femmes	932	1661	2593
%	35,9	64,1	100

Total	1677	3320	4997
%	33,6	66,4	100

La lecture rapide qui consiste à repérer le signe des écarts à l'indépendance (par le biais des écarts au pourcentage moyen) nous donne la structure suivante :

S'occupe d'un jardin d'agrément		
	Oui	Non
Hommes	-	+
Femmes	+	-

On voit que les femmes sont plus nombreuses que les hommes à s'occuper d'un jardin d'agrément, pratique qui est le fait d'un tiers de la population. Le jardin d'agrément est marqué féminin.

Inversement on a les résultats suivants pour le jardin potager :

S'occupe d'un jardin potager			
	Oui	Non	Total
Hommes	671	1733	2404
%	27,9	72,1	100
Femmes	520	2073	2593
%	20,1	79,9	100

Total	1191	3806	4997
%	23,8	76,2	100

La situation est inverse : la lecture rapide nous donne la structure suivante :

	S'occupe d'un jardin potager	
	Oui	Non
Hommes	+	-
Femmes	-	+

Le jardin potager est marqué masculin avec tous les stéréotypes associés : nécessité de la force physique, production utile, alors que les stéréotypes associés au jardin d'agrément sont "la grâce et la délicatesse féminine" associées à la cueillette des roses. Le marquage social est moins fort que dans le cas de la chasse ou de la fabrication des vêtements, il existe cependant, sauf quand on envisage le mélange des deux activités où les tendances vers les attractions féminines ou masculines s'annulent.¹⁵

Résumons les outils mis au point : *l'écart au pourcentage moyen* nous permet de rapidement repérer le signe des *écarts à l'indépendance* (attraction ou répulsion). Il ne faut pas confondre ces deux outils : l'écart au pourcentage moyen est en général calculé en ligne (c'est la différence entre le pourcentage de la ligne et le pourcentage toutes lignes confondues) : c'est donc une *différence de pourcentage*.

L'écart à l'indépendance est un *effectif* et c'est un invariant, indépendant du choix des lignes et des colonnes (c'est la différence entre l'effectif observé et l'effectif théorique : le résultat est donc un effectif). Seuls les signes sont les mêmes pour ces deux outils et c'est ce qui permet d'utiliser l'écart au pourcentage moyen, particulièrement

¹⁵ Le tableau initial (jardinage sans autre indication) n'est pas la somme des deux tableaux spécifiques (potager et agrément), mais leur mélange qui ne peut être qu'observé et non calculé car 851 personnes pratiquent les deux types de jardinages.

simple à calculer, pour arriver facilement à une lecture rapide du tableau où l'on ne considère et interprète que les signes des écarts à l'indépendance.

Une lecture rapide est un instrument qui a les qualités et les limites des techniques analogues de lecture des livres : elle permet de se faire une idée approchée de la structure des écarts d'un tableau croisé. Sa rapidité est son avantage mais elle doit souvent être complétée par une lecture exhaustive de l'ensemble des écarts du tableau. C'est ce qui va être rendu possible par la technique de lecture intégrale d'un tableau par le biais de sa visualisation graphique.

4 La visualisation de l'écart à l'indépendance

Faire un graphique, c'est tirer d'un ensemble de nombres ceux que l'on juge pertinents de représenter et, à cette fin, on remplace en général ces nombres par des longueurs (diagrammes en bâtonnets) ou par des surfaces (diagrammes en part de *gâteaux*¹⁶). Ces visualisations graphiques sont efficaces pour des distributions simples (même si elles sont nombreuses) mais non pour des tableaux croisés dont on a vu que l'élément utile et invariant était l'écart à l'indépendance.

Avant de représenter ces écarts, nous allons montrer comment on peut y arriver en commençant par représenter, non pas les écarts, mais l'ensemble des effectifs du tableau par des surfaces proportionnelles aux effectifs de chaque case du tableau. Nous travaillerons sur un tableau plus vaste que les précédents : il sera à 3 lignes et 4 colonnes.

Les données sont tirées de l'enquête LATIN94¹⁷. De cette enquête faite auprès de 1058 répondants de la région parisienne de classe

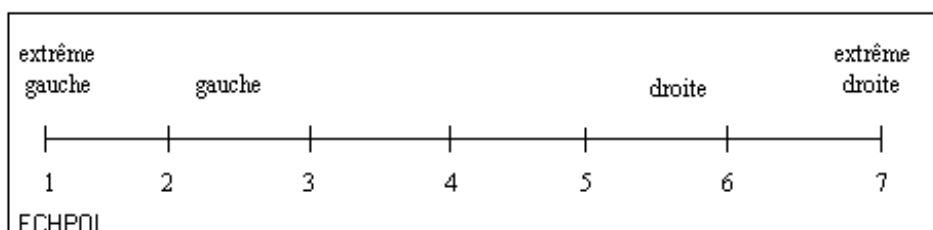
¹⁶ ou de *camembert* selon les cultures nationales

¹⁷ Enquête faite en 1994 dans le cadre de l'enseignement de méthodologie du Deug de Paris V en collaboration avec François de Singly. Elle est centrée sur

moyenne ou supérieure (ni ouvriers, ni employés) parents d'un enfant en Cours moyen à l'école primaire, on tire les deux questions suivantes

1) proximité politique : "On classe habituellement les gens sur une échelle de ce genre où, comme vous le voyez, on peut classer les gens plus ou moins à gauche ou plus au moins à droite. Vous personnellement, pouvez-vous me dire, par un numéro, où vous vous situez sur cette échelle ?" (montrer échelle politique)

- | | |
|-------------------|-------------------|
| 1. extrême gauche | 5. centre droit |
| 2. gauche | 6. droite |
| 3. centre gauche | 7. extrême droite |
| 4. centre | (0. Non réponse) |



l'enseignement du latin : on a fait l'hypothèse que l'attitude vis à vis de cet enseignement pouvait avoir des liens avec les opinions politiques et religieuses.

Cette question en 7 modalités est recodée¹⁸ de la façon suivante :

- 1) gauche (anciennes modalités 1 à 3)
- 2) centre (anciennes modalités 4 ou non réponse 0)
- 3) droite (anciennes modalités 5 à 7)

2) positionnement religieux :

Vous considérez-vous comme :

- | | |
|---|---------------------------|
| 1. catholique pratiquant | 5. protestant |
| 2. catholique non pratiquant | 6. juif |
| 3. de tradition catholique mais
ni croyant ni pratiquant | 7. autre religion (noter) |
| 4. sans religion | (0. Non réponse) |

On élimine les individus ayant pris les modalités 5 à 7 et 0 qui correspondent à des effectifs faibles et qui demanderaient une enquête spécifique et l'on obtient le tableau croisé suivant :

		Choix religieux			Total
Choix pol.	Pratiquant	Non prat	Tradition	Sans religion	
Gauche	34	95	170	149	448
Centre	51	113	93	52	309

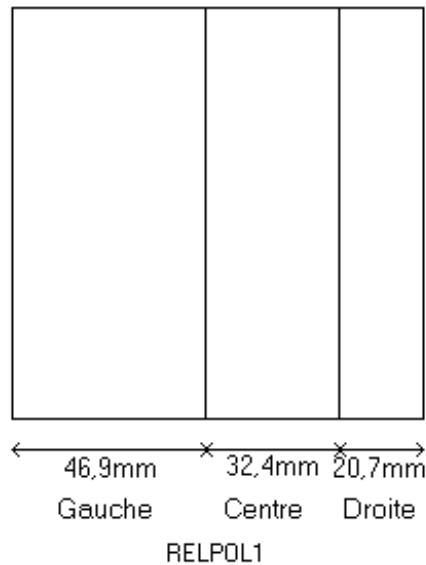
¹⁸ recoder signifie agréger dans un même code ce qui était auparavant dans des modalités différentes : ces opérations sont faites d'une manière logicielle (ici par l'intermédiaire du programme OUTILS du logiciel TRIDEUX)

Droite	61	85	33	19	198
Total	146	293	296	220	955

on calcule les pourcentages en ligne et des marges :

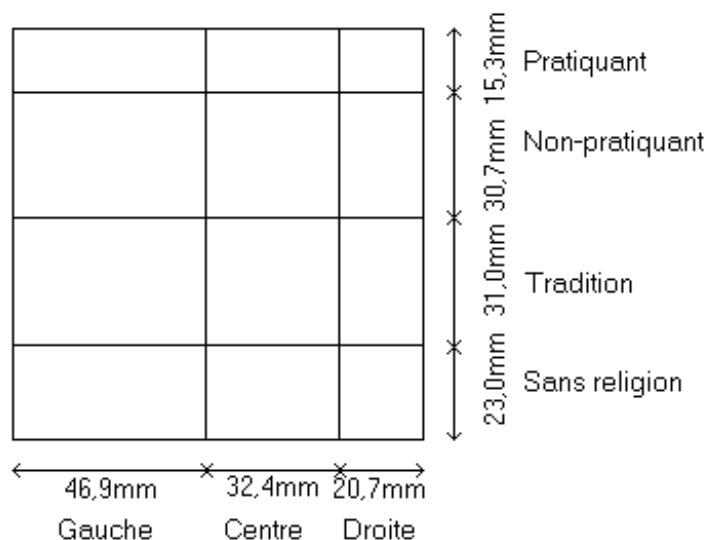
Choix pol.	Pratiquant	Non prat	Tradition	Sans religion	Total
Gauche	7,6	21,2	37,9	33,3	100 46,9
Centre	16,5	36,6	30,1	16,8	100 32,4
Droite	30,8	42,9	16,7	9,6	100 20,7
Total	15,3	30,7	31,0	23,0	100 100

Pour faire une représentation graphique de l'ensemble, il faut déjà constater que la gauche, qui représente 46,9% de la population (marge du total en colonne) doit représenter 46,9% de la surface du graphique, le centre 32,4% et la droite 20,7%. A cette fin on considère une surface carrée (pour des essais sur papier millimétré par exemple de 10 cm sur 10 cm) que l'on divise en trois bandes, la première de gauche de 46,9 mm (arrondis à 47 mm), puis de 32 mm pour le centre et 21 mm pour la droite. Soit le graphique RELPOL1 qui respecte ces proportions (mais non ces unités).



Puisque l'ensemble de la surface correspond à 100% de la population, la première bande qui représente 46,9% de la largeur correspond à cette proportion de la population et représente donc la population qui se classe politiquement à gauche. On a de même les deux autres bandes proportionnelles au Centre et à la Droite.

Divisons maintenant l'autre dimension de la surface en tenant compte de l'autre point de vue, celui de la pratique religieuse. On peut diviser l'autre dimension proportionnellement aux pourcentages marginaux comme ci-dessous dans le graphique RELPOL2:



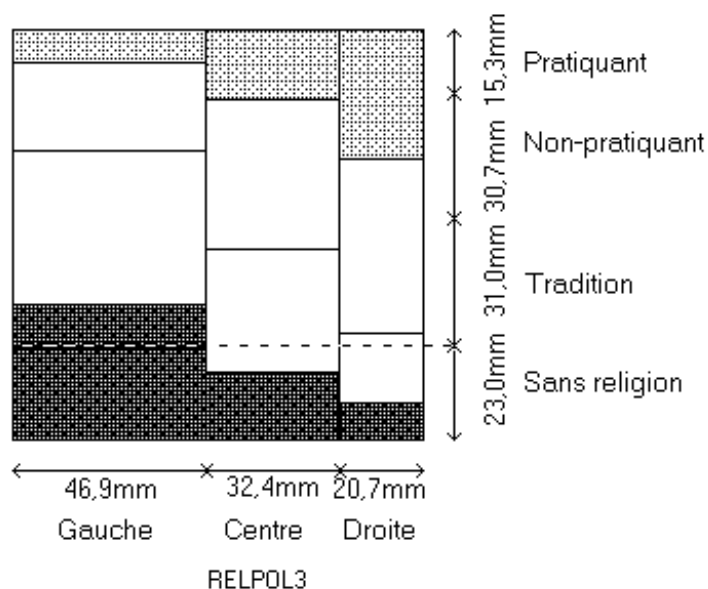
RELPOL2

Nous avons bien divisé la surface totale proportionnellement aux pourcentages des marges mais le résultat obtenu n'est pas la représentation du tableau croisé mais une réalité qui ne dépend que des marges, c'est à dire un tableau proportionnel aux effectifs théoriques. En effet, soit la surface de la première case du tableau, les pratiquants de gauche. De la bande de la Gauche (qui correspond à 46,9% de la surface), on prend les 15,3% qui correspondent aux pratiquants. Pour faciliter les calculs on utilise les proportions correspondantes : prendre 15,3% de 46,9% cela revient à multiplier 0,153 par 0,469 soit en revenant aux effectifs qui permettent de calculer ces proportions $146 / 955 \times 448 / 955 = 0,0717$. Cette case des pratiquants de gauche correspond à 7,2% de la population. Pour retrouver l'effectif correspondant, il suffit de multiplier la proportion par l'effectif total soit $0,0717 \times 955 = 68,5$ individus. En multipliant la proportion avant calcul ($146 / 955 \times 448 / 955$) par l'effectif, il y a une simplification par l'effectif 955 qui se retrouve une fois en bas et une fois en haut et

l'on retrouve le résultat :

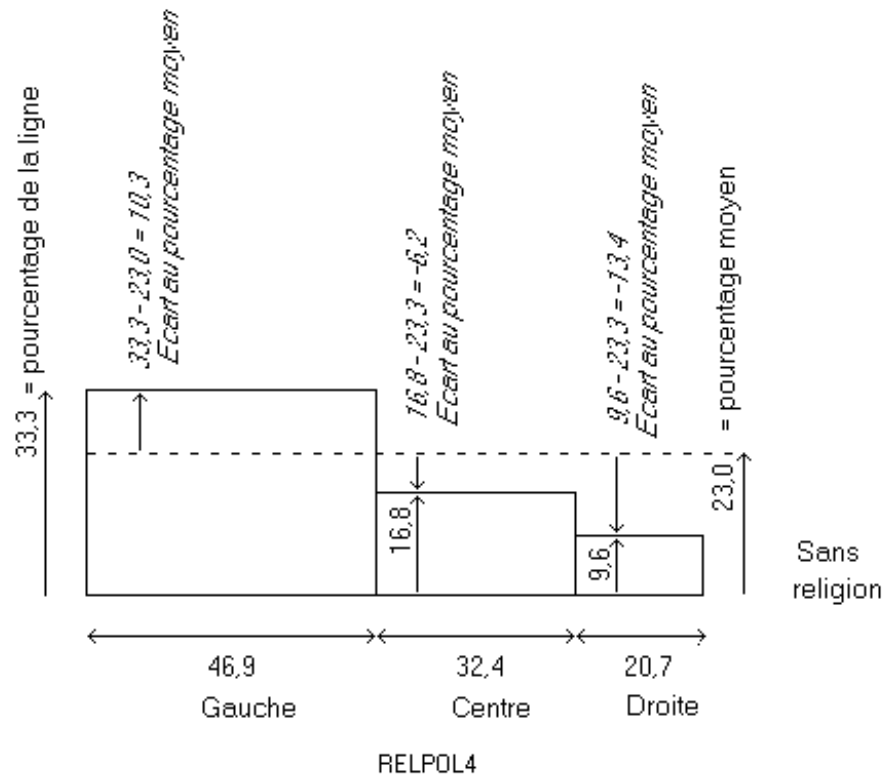
$146 \times 448 / 955 = 68,5$ où l'on reconnaît l'effectif théorique, produit des marges par le total.

En conclusion, une surface quelconque du tableau est proportionnelle à l'effectif théorique de la case considérée. Ce résultat est utilisé par un certain nombre de représentations graphiques qui visualisent dans la case l'intensité de l'attraction ou de la répulsion, mais le graphique ne visualise toujours pas les écarts à l'indépendance. La raison en est simple, c'est que les différentes bandes de la surface (qui correspondent aux lignes du tableau), doivent avoir chacune leur division propre exprimée par le pourcentage de chaque ligne. On va donc dans chaque bande faire la division selon la répartition de chaque ligne.



Dans le graphique RELPOL3, chaque bande est divisée selon le pourcentage en ligne : par exemple les sans religion de gauche représentent 33,3% de la gauche, ce qui est davantage que le pourcentage moyen de 23,0% que l'on a laissé comme repère à droite.

On a tramé les surfaces correspondant aux cases extrêmes du tableau :
par exemple, faisons un grossissement pour les *sans religion* :



Le pourcentage dans la gauche est de 33,3% (pourcentage en ligne), donc plus que la moyenne qui est de 23,0% (pourcentage toutes lignes confondues). Au centre le pourcentage est de 16,8% et à droite de 9,6%.

Dans ces trois bandes, on peut repérer les surfaces qui correspondent à des écarts au pourcentage moyen : dans la bande de gauche, c'est la surface de largeur 46,9mm qui dépasse la hauteur moyenne de 23,0mm (*écart de 10,3mm*). Dans la bande du centre, c'est la surface de largeur 32,4mm comprise entre la hauteur de la case (à 16,8mm) et la hauteur moyenne de 23,0 ; de hauteur égale à la différence entre le pourcentage de la ligne et le pourcentage moyen

toutes lignes confondues $16,8 - 23,0 = - 6,2$: le résultat est négatif car c'est un déficit par rapport à la moyenne (le pourcentage moyen). Enfin pour la droite, la largeur est de 20,7mm et la hauteur égale à $9,6 - 23,0 = -13,4$

Vérifions : la base du premier rectangle est ce que nous appellerons l'importance globale, ou le *poids* de la gauche. En pourcentage c'est 46,9% du total, en proportion c'est plus précisément 448 personnes de gauche sur 955 : $448 / 955 = 0,4691$.

PoidsGauche = 0,4691 et de même

PoidsCentre = 0,3236

PoidsDroite = 0,2073

la somme des poids est l'unité (à l'arrondi près éventuellement).

La hauteur du premier rectangle, c'est la différence entre le pourcentage de la ligne 33,3 et le pourcentage toutes lignes confondues, 23,0. En proportion, appelons Proportion de la gauche pour les sans religion (en abrégé PropGaucheSR) le rapport $149 / 448 = 0,3326$

PropGaucheSR = 0,3326 et de même

PropCentreSR = 0,1683

PpropDroiteSR = 0,0960

la proportion des sans religion toute lignes confondues est :

PropSR = 0,2304

La hauteur du premier rectangle est égale à $\text{PropGaucheSR} - \text{PropSR} = 0,1022$ c'est l'écart au pourcentage moyen que nous nommons EcartGaucheSR.

La surface est le produit de la largeur par la hauteur, produit de PoidsGauche par EcartGaucheSR soit $0,4691 \times 0,1022 = 0,04794$.

Cette proportion d'un effectif total de 955 (appelé Total) correspond à un effectif de

$0,04794 \times \text{Total}$ soit $0,04794 \times 955 = 45,8$ individus. Vérifions que c'est bien l'effectif en écart à l'indépendance :

L'effectif théorique produit des marges par le total :

$$448 \times 220 / 955 = 103,2$$

L'effectif observé est de 149 ;

observé - théorique = $149 - 103,2 = 45,8$ on retrouve bien la même valeur correspondant à l'écart à l'indépendance.

Pour les deux autres cases on a en proportion :

$$\text{- EcartCentreSR} = \text{PoidsCentre} \times (\text{PropCentreSR} - \text{PropSR}) = -0,02010$$

$$\text{et en effectif EcartCentreSR} \times \text{Total} = -0,02010 \times 955 = -19,2$$

$$\text{- EcartDroiteSR} = \text{PoidsDroite} \times (\text{PropDroiteSR} - \text{PropSR}) = -0,02786$$

$$\text{et en effectif EcartDroiteSR} \times \text{Total} = -0,02786 \times 955 = -26,6$$

En résumé :

$$\text{Effectif EcartGaucheSR} = 45,8$$

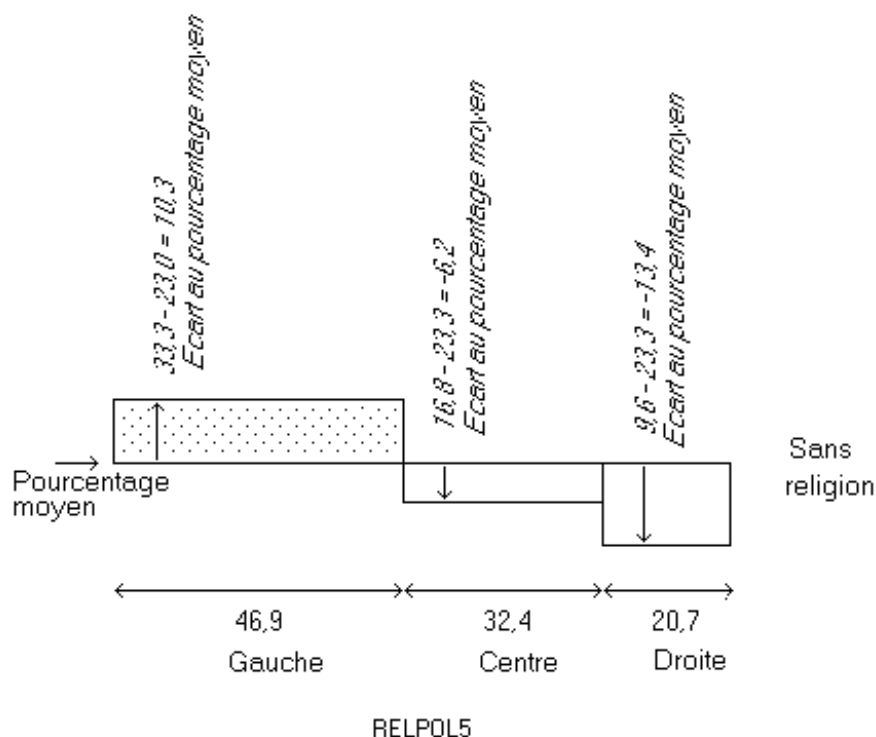
$$\text{Effectif EcartCentreSR} = -19,2$$

$$\text{Effectif EcartDroiteSR} = -26,6$$

$$\text{Somme des écarts négatifs} = -45,8$$

On voit immédiatement que la surface en écart positif (au-dessus) est égale à la somme des écarts négatifs (en-dessous)

Pour visualiser les écarts à l'indépendance, il suffit donc de gommer le bas du graphique et de ne retenir que les surfaces correspondants aux écarts à l'indépendance : au-dessus de la moyenne pour la gauche, en-dessous pour le centre et la droite. On a maintenant le graphique RELPOL5 suivant :



Pour mieux faire apparaître les attractions qui correspondent aux écarts positifs à l'indépendance, on trame les surfaces correspondantes. Il reste à obtenir les profils correspondant aux autres colonnes du tableau. Il n'est pas nécessaire de repasser par toutes les étapes précédentes puisqu'on constate qu'il suffit de prendre en compte les écarts au pourcentage moyen de chaque colonne et, sur un trait horizontal affecté au pourcentage moyen, de construire des rectangles, de mêmes bases pour tous les profils, au-dessus ou en-dessous du pourcentage moyen selon le signe de l'écart, et de hauteur égal à l'écart au pourcentage moyen.

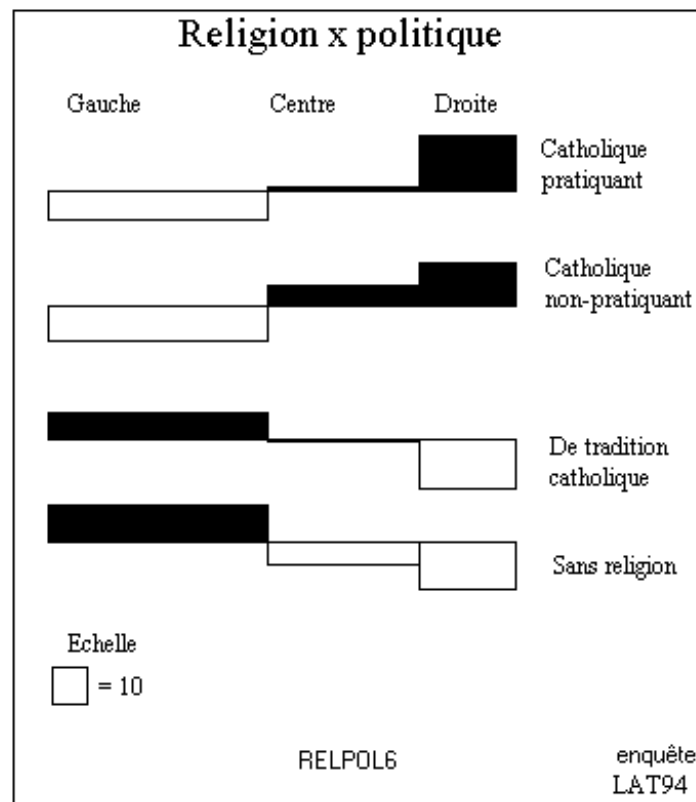
Soit le calcul des écarts au pourcentage moyen :

	Pratiquant	Non prat	Tradition	Sans religion
Gauche	-7,7	-9,5	+6,9	+10,3
Centre	+1,2	+5,9	-0,9	-6,2
Droite	+15,5	+12,2	-14,3	-13,4

A titre d'exemple, l'écart positif qui marque l'attraction entre « Droite » et « pratiquant » à pour base une longueur proportionnelle à 20,7, pourcentage de la ligne « droite » et une hauteur positive proportionnelle à +15,5, écart au pourcentage moyen de cette case.

Finalement après avoir fait la même opération pour les autres colonnes, et avoir ajouté une échelle¹⁹, on a la représentation en surface des écarts à l'indépendance du suivante :

¹⁹ Echelle obtenue de la manière suivante : si le carré initial de 100mm sur 100mm soit 10.000mm² correspond au total de 955 individus, à 1 individus correspond une surface de 10.000/955 mm² = 10,47 mm² et donc à un carré de côté égal à la racine carrée de 10,47 soit 3,2mm. Pour 10 individus la surface est de 104,7mm² et le côté du rectangle est de 10,2mm ce qui est tout à fait lisible.



Interprétation

Le but d'une représentation graphique est de permettre une lecture simultanée de tous les éléments pertinents du tableau (attractions notées par une surface au-dessus et soulignés en noir car c'est ce qui est le plus intéressant ; absence d'écarts ou indépendance notée par une simple ligne horizontale ; oppositions notées par une surface en dessous).

On voit aisément :

- que catholiques pratiquants et non-pratiquants sont plutôt en attraction avec la droite ;

- que sans religion et personnes de tradition catholique se situent plutôt en attraction avec la gauche.

Il s'agit là d'un phénomène bien connu²⁰ et lié à l'histoire des rapports entre catholicisme et vie politique en France depuis la Révolution. En effet la Révolution dans ses débuts n'a pas été antireligieuse : le clergé a participé aux cahiers de doléances et la fête de la fédération du 14 juillet 1790 a connu sa Grand-messe (célébrée par Talleyrand). Ce n'est qu'au moment où les révolutionnaires, semble-t-il essentiellement pour des raisons financières, ont voulu pousser à bout les tendances gallicanes de l'Eglise de France, c'est à dire faire une Eglise de France qui soit autonome vis-à-vis de Rome, qu'une partie du clergé s'est opposée à cette rupture et a refusé de jurer fidélité au nouveau régime. Prêtres jureurs (soutenus par le pouvoir) et réfractaires (recherchés et persécutés) ont divisé la société d'une façon profonde. Le clergé réfractaire est resté dans un même geste fidèle à Rome et fidèle au Roi. De l'autre côté, le radicalisme antireligieux de la Révolution s'appuie sur cette fidélité pour voir dans le catholicisme un soutien de l'étranger qui veut la mort de l'idée révolutionnaire.

Napoléon calme le jeu par le Concordat de 1801 qui durera un siècle pendant lequel l'amalgame créé à la Révolution va se maintenir : les forces qui ont imposé la Restauration sont liées au catholicisme ;

²⁰cf. sur l'aspect politique : René Rémond, *La droite en France*, Paris, Aubier, 1963 ; Jean-François Sirinelli (dir.) *Histoire des droites en France*, Paris, Gallimard, 1992, 3 volumes (repris partiellement en poche sous le titre : *Les droites françaises*, folio histoire n°63) ; sur la question scolaire, Alain Prost, *L'enseignement en France 1800-1967*, Paris, Armand Colin, 1968 ; sur les rapports spécifiques entre classe sociale et religion cf. Guy Michelat et Michel Simon, *Classe religion et comportement politique*, Paris, Presses de la FNSP et ed. sociales, 1977 et la nouvelle appréciation de : Jean-Marie Donegani, *La liberté de choisir*, Paris, Presses de la FNSP, 1993.

les forces qui luttent pour continuer l'œuvre de la Révolution s'opposent à la volonté politique de l'Eglise catholique et luttent contre elle pour cette raison. Sous la troisième République, l'opposition est forte entre gauche et droite de l'époque : entre le camp républicain et le camp légitimiste. Progressivement, les républicains vont vouloir affranchir la politique française de la domination de l'Eglise catholique, en particulier en ce qui concerne l'école car ils ont bien conscience que l'emprise sur les esprits passe par ce biais. Bien que les catholiques aient fini par se rallier à la République, soutenu dans cette attitude par le Pape Léon XIII, cette tendance à la dissociation trouvera son point culminant avec la Séparation de l'Eglise et de l'Etat de 1904.

Depuis, l'opposition sur la question scolaire s'est amalgamée à l'opposition gauche/droite. Il est symptomatique de voir que les manifestations les plus importantes des années 1980 et 1990 ont été liées à la question scolaire : manifestations de plus d'un million de personnes pour la défense de l'école libre (c'est à dire essentiellement catholique) quand les socialistes se sont installés au pouvoir en 81 et ont voulu créer une école unique qui intégrerait toutes les écoles touchant des fonds publics. Inversement la gauche a fait une manifestation analogue par son ampleur quand, sous un gouvernement de cohabitation de droite, une tentative de modification de la loi Falloux a été envisagée (cette loi de compromis faite au milieu du 19^e siècle n'autorise pas actuellement un financement trop important des écoles par les collectivités locales)

On peut donc dire que les attractions entre gauche et opposition à l'Eglise, et droite et adhésion au catholicisme, que l'on observe sur le graphique RELPOL6 sont en quelque sorte le résultat d'une secousse sociale vieille de deux siècles : on a célébré il y a quelques années le bicentenaire de la Révolution mais sans vraiment que les gens se sentent personnellement concernés alors que, paradoxalement, beaucoup maintiennent fermement cette association ancienne entre

politique et religion. Cette association, assez spécifique de l'histoire de France (ou des pays latins), est peu vécue comme telle dans les pays anglo-saxons. Ce tri croisé est la manifestation d'une conjecture politique qui se maintient en l'état, décennies après décennies, d'une manière analogue à ces enregistrements du climat passé que nous restituent les cernes de croissances des arbres et qui permettent de dater un tronc d'arbre vieux de plusieurs siècles.

On peut aller plus loin et faire l'hypothèse que se dire « sans religion », ce n'est pas être sans religion, c'est prendre ses distances vis-à-vis de la religion dans laquelle on est né. Cette hypothèse trouve sa confirmation par exemple dans l'enquête REL86²¹ faite auprès d'un échantillon représentatif de 1530 personnes de nationalité française : si pratiquement tous les catholiques se disent baptisés, cette proportion est encore de 83% pour ceux qui se sont déclarés « sans religion », ce qui manifeste bien que la plupart d'entre eux sont nés dans la religion catholique, et que s'ils se déclarent sans religion, c'est qu'ils veulent prendre leurs distances vis-à-vis de l'Eglise catholique et que pour cela ils se situent à gauche²².

L'apport propre de ce tableau par rapport à cette problématique se trouve dans le parallélisme entre les deux profils « de tradition catholique » et « sans religion » avec cette nuance que ceux qui ont encore un rapport avec le catholicisme, rapport assez extérieur puisqu'ils se définissent comme ni pratiquant et ni croyant, sont plus enclins que les sans religion à choisir le centre (avec lequel ils sont en

²¹ Guy Michelat, Julien Potel, Jacques Sutter, Jacques Maître, *Les Français sont-ils encore catholiques ?*, Paris, Cerf, 1991, rend compte d'une enquête réalisée par ces chercheurs en 1986 désignée dans la suite par le sigle REL86

²² on raisonne ici en termes d'attractions : il est évident qu'il existe des catholiques de gauche et ils sont présents dans le tableau croisé, comme ils sont présents socialement dans divers institutions (à la CFDT, au PS, dans une revue comme *Esprit*)

situation d'indépendance et non d'opposition). Leur position est moins extrême. D'autre part, et de la même manière que les sans religion sont plus à gauche que les catholiques de tradition, les catholiques pratiquants sont plus à droite que les non pratiquants où le centre est plus marqué.

Ce qui est aussi intéressant, c'est de noter que les répondants « tradition culturelle » sont les plus nombreux : des 4 catégories d'appartenance, c'est la plus importante avec 31%. Ce choix manifeste que proposer cette réponse (ce qui n'était pas fait jusqu'à présent dans les enquêtes de sociologie religieuse), était cohérent avec les attentes des répondants. En effet, ce choix peut sembler logiquement critiquable car qu'est-ce qu'être catholique quand on n'est ni croyant, ni pratiquant ? Il faut bien voir que refuser le concept de "ni croyant ni pratiquant", c'est se situer dans le système de croyance catholique où cette position de catholique de tradition culturelle est négatrice de l'identité catholique de l'institution. D'ailleurs, il est plus facile pour l'institution catholique de penser les sans religion comme tels, plutôt que comme des catholiques en opposition avec l'Eglise. Pour une institution quelconque, rejeter ses contestataires à l'extérieur est plus facile que de les prendre en compte comme contestataires. On peut penser que la lente érosion de la pratique catholique (30% de pratiquants après la guerre contre 10% aujourd'hui²³) est un symptôme de cette prise de distance, de cette position « d'opposants concernés » qui se sentent solidaires d'un passé sans se retrouver dans le discours de l'institution présente²⁴. Par leur encrage à gauche, par leur acceptation d'être considéré comme catholiques, ces « tradition culturelle » sont en position intermédiaire entre les catholiques qui

²³ Guy Michelat *et al.*, *Les Français ...* (REL86), p.35

²⁴ cette existence d'un « christianisme culturel » est bien mise en valeur par Yves Lambert dans « La religion : paysage en pleine évolution » in Hélène Riffault (dir.), *Les valeurs des français*, Paris, PUF, 1994.

sont encore dans l'emprise culturelle de l'Eglise (les croyant non pratiquants) et ceux qui ne se sentent plus aucun lien avec l'Eglise (les sans religion).

La conclusion de cette interprétation est que l'on peut donner un sens complémentaire à l'intitulé de la question d'appartenance religieuse : on peut l'interpréter, *dans toutes ses modalités*, comme des degrés d'une échelle d'appartenance ou de rejet de l'institution catholique en France. Paradoxalement, la réponse permet de mieux comprendre la question : c'est souvent le cas car interpréter, c'est appliquer à la réalité sociale une grille de lecture (une problématique, une théorie), qui modifie une perception ordinaire ou la perception d'une autre problématique.

Dans le cas présent, l'ancienne problématique est celle d'un courant de la sociologie religieuse qui est né à l'intérieur de l'Eglise catholique et dont le but était de se rendre compte objectivement de la réalité de la pratique à des fins de meilleure gestion du clergé. Il fallait alors compter les gens en fonction de leur plus ou moins grand respect des règles édictées par l'Eglise : accepter de distinguer ceux qui pratiquaient régulièrement de ceux qui avaient une pratique occasionnelle (grandes fêtes ou événements sociaux comme mariages ou enterrements) était déjà un progrès théorique dans la mesure où cette sociologie religieuse admettait a priori que l'on prenne ses distances avec l'obligation stricte de la pratique chaque dimanche.

La nouvelle problématique est issue du fait qu'une nouvelle distance peut être prise vis-à-vis de l'institution religieuse en proposant la possibilité d'une appartenance catholique qui ne se reconnaisse pas dans le discours d'aujourd'hui de l'institution sans pour autant se sentir à l'extérieur. Du point de vue théorique, ce qui est paradoxal, c'est que tant qu'on pose l'ancien style de question (pratiquant, non pratiquant, sans religion), on en peut pas voir ce qui relève de la nouvelle problématique et on peut donc dire qu'elle n'a

aucune raison d'être ; alors que dès qu'on pose la nouvelle question, la force du groupe « tradition culturelle » impose du même fait cette problématique. On trouve sur cette exemple vérification du fait que l'on ne passe pas progressivement d'une problématique à une autre, mais par des sauts²⁵. Dans l'ancienne problématique, des difficultés apparaissaient car si l'on étudie les croyances de ceux qui se déclarent catholique, on constate que leurs valeurs sont situées en référence avec d'autres sources que l'Eglise²⁶. Pour résoudre ces problèmes, une nouvelle problématique entraînant de nouvelles manière de poser les questions a vu le jour et c'est celle-ci dont on trouve une vérification dans le présent tableau croisé.

5 Exemples de représentation en surface

Mobilité sociale

Soit le tableau suivant qui croise, pour une population de 6.022.000 hommes âgés de 40 à 59 ans en 1993 ayant (ou ayant eu) un emploi, la catégorie socioprofessionnelle (CSP) qui est la leur au moment de l'enquête (colonnes du tableau) et la CSP de leur père (en ligne). On donne les effectifs (en milliers), les pourcentages en ligne et les pourcentages des marges²⁷. Les pourcentages en lignes supérieurs au

²⁵ Thomas S. Kuhn, *La structure des révolutions scientifiques*, Paris, Flammarion, 1983, traduction de la nouvelle édition de 1970 (première édition 1962)

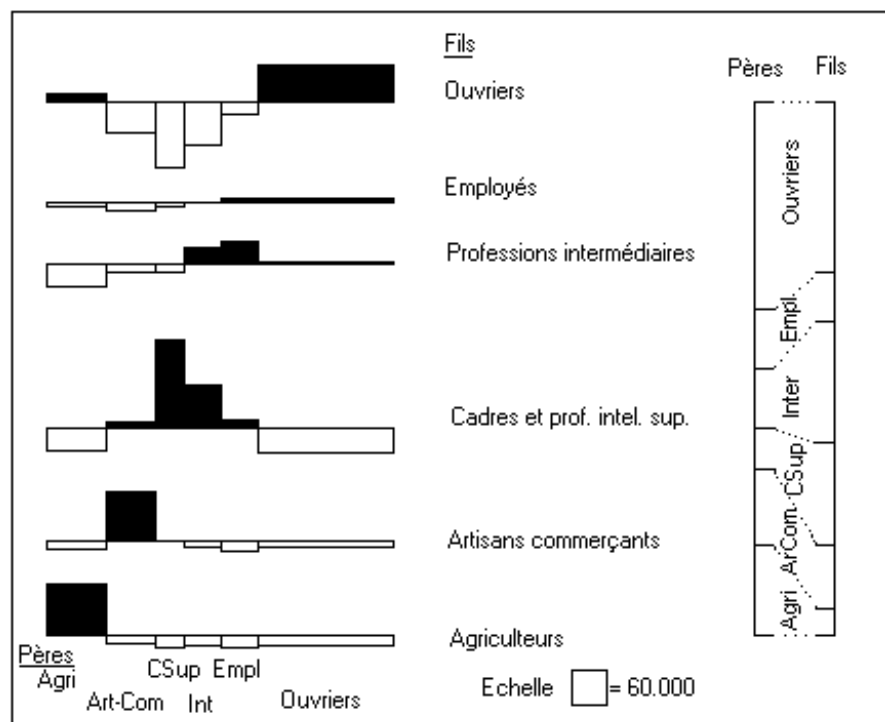
²⁶ Lambert 1994, p.136

²⁷ Ce tableau, issu de l'enquête FQP 1993 de l'Insee (Enquête sur la Formation et la Qualification Professionnelle), est donné dans le recueil de l'Insee *Données sociales 1996*, p.311. Les CSP utilisées sont les suivantes Agriculteurs, Artisans commerçants et chefs d'entreprise (ArCo), Cadres et professions intellectuelles supérieures (CSup), Professions intermédiaires, Employés, Ouvriers.

pourcentage moyen, et qui correspondent donc à des attractions, ont été mis en gras pour permettre une lecture rapide.

CSP du père	Catégorie socioprofessionnelle du fils en 1993						Total
	Agric	ArCo	CSup	Inter	Empl	Ouvr	
Agric	258 24.6	81 7.7	108 10.3	153 14.6	84 8.0	365 34.8	1049 (17.4) 100
ArCo	14 1.7	246 29.6	180 21.7	168 20.2	56 6.7	167 20.1	831 (13.8) 100
CSup	3 0.6	54 10.7	266 52.9	104 20.7	42 8.3	34 6.8	503 (8.4) 100
Inter	5 0.8	56 8.8	225 35.5	190 30.0	61 9.6	97 15.3	634 (10.5) 100
Empl	1 0.1	49 7.3	148 22.2	215 32.2	74 11.1	180 27.0	667 (11.1) 100
Ouvr	19 0.8	204 8.7	228 9.8	568 24.3	251 10.7	1068 45.7	2338 (38.8) 100
Total	300 5.0	690 11.5	1155 19.2	1398 23.2	568 9.4	1911 31.7	6022 (100) 100

On représente en surface les écarts à l'indépendance dans le graphique MOB93 :



Catégorie socioprofessionnelle des fils en 1993 en fonction de celle des pères

MOB93

En plus de la représentation en surface des écarts à l'indépendance, on trouve dans ce graphique MOB93 une échelle et, à droite, une représentation des pourcentages des marges du tableau afin de pouvoir repérer visuellement l'évolution des différentes CSP entre la génération des pères et celle des fils. On repère à partir de cette évolution les changements profonds subis par la structure sociale entre ces deux générations.

En effet puisqu'il s'agit d'hommes de 40 à 59 ans, leurs pères, au moment de leur naissance, étaient actifs des années 1935 aux années 1955, donc autour de la 2^e guerre mondiale. On voit donc d'abord la trace de l'exode rural : la CSP des agriculteurs représentait 17% des

pères, elle n'en représente plus que 5% pour les fils. Beaucoup d'actifs ont donc dû, entre les deux générations, quitter l'agriculture. La proportion des ouvriers a également diminué passant de 39 à 32%. Artisans-commerçants et employés étant restés à peu près stables, les augmentations se situent au niveau des CSP restantes : professions intermédiaires (CSP qui passe de 11 à 23%) et cadres et professions intellectuelles supérieures (8 à 19%).

Ces évolutions, liées au développement d'un salariat ouvrier (au détriment de l'agriculture), puis à sa réduction au profit d'emplois plus qualifiés, nous permettent maintenant d'interpréter le contenu du tableau en examinant les différents profils qui correspondent aux CSP des fils.

Un profil est très plat : celui des employés. Ceci signifie qu'il y a peu d'écarts à l'indépendance : la CSP *employé* représente 9% du total et c'est ce pourcentage que l'on retrouve à quelques points près dans les pourcentages en ligne de cette colonne du tableau (le plus bas correspond à un déficit d'Artisans-commerçants de 2,7%, le plus haut à une attraction avec l'origine sociale du père identique, employé, à +1,7%). Ceci manifeste que cette catégorie est alimentée d'une manière équivalente par toutes les autres catégories bien qu'il y ait attraction faible avec l'origine ouvrière ou l'origine employé et répulsion faible avec les autres CSP. Par contre, si on considère, non plus le profil employé (fils) mais l'empilement vertical des différents rectangles employés (pères) dans les autres profils, on repère deux attractions (en plus de celle employé père et fils), c'est à dire des approvisionnements préférentiels dans les profils *Cadres* et *Intermédiaires*. Dans cette dernière catégorie, le rectangle *employé* est même de niveau supérieur au niveau *intermédiaire* (père et fils en ligne et en colonne), ce qui est le seul cas analogue dans tout le

tableau où le phénomène de continuation *tel père, tel fils*²⁸ (correspondant aux cases diagonales du tableau), est toujours le plus fort. La catégorie d'employé est un sorte de plaque tournante : on y vient plutôt venant du monde ouvrier, on en sort plutôt vers les professions intermédiaires et vers les cadres.

Le même phénomène se retrouve avec les *professions intermédiaires* qui n'ont pas encore un profil aussi marqué que les autres CSP. Ces professions, "intermédiaires" comme leur nom le suggère, servent aussi d'étape entre le monde employé qui les alimente particulièrement et les cadres qu'elles alimentent.

Les autres profils sont spécifiques : les agriculteurs sont en priorité des fils d'agriculteurs. Les ouvriers sont issus particulièrement d'ouvriers mais aussi d'agriculteurs. Les artisans-commerçants se recrutent en leur sein propre. Reste le profil des cadres et professions intellectuelles supérieures : c'est la seule catégorie qui s'alimente à quatre CSP : auprès d'elle-même, d'une manière privilégiée (case diagonale) mais aussi auprès des professions intermédiaires, des employés et des artisans-commerçants. Cette CSP a forte croissance a ratissé large : de nombreuses opportunités ont été proposées à des jeunes de nombreuses origines.

Si l'on regarde l'ensemble du graphique, les grandes masses noires s'inscrivent dans une logique *tel père, tel fils*. Il y a beaucoup de continuité sociale d'une génération à l'autre mais ce qui a frappé cette génération née aux alentours de la 2^e guerre mondiale et qui a connu une croissance économique importante jusqu'aux années 70, c'est la perception qu'il y avait des opportunités d'ascension sociale. Beaucoup

²⁸ Cf. Claude Thélot, *Tel Père, tels fils ? Origine familiale et position sociale*, Paris, Dunod, 1982. On trouvera une bibliographie d'ensemble dans Dominique Merllié, *Les enquêtes de mobilité sociale*, Paris, Presses Universitaires de France, 1994

de pères de cette génération qui étaient dans une position basse à leurs yeux, ont pensé, souvent avec raison, que leur fils pouvait avoir une position sociale plus élevée que la leur (et que leur fille pourrait peut-être arriver au même résultat par le biais d'un mariage avec quelqu'un en montée sociale), la clé de l'ascension sociale se faisant par l'école. Ces cas de mobilité sociale se trouvent dans les cases hors diagonales en écart positif, importantes dans les profils cadres et professions intermédiaires.

On soulignera un aspect technique en conclusion de cette analyse : en représentant les colonnes du tableau comme des profils, on fait le choix de porter l'attention davantage sur les colonnes que sur les lignes. En inversant lignes et colonnes, les nouveaux profils représenteraient les *mêmes surfaces* en écart à l'indépendance (puisque les écarts à l'indépendance sont des invariants) mais de façon différente, ni les largeurs ni les hauteurs ne seraient les mêmes bien que leur produit corresponde aux mêmes surfaces.

Cependant, on a vu que l'on peut "prendre le point de vue des lignes" simplement en envisageant les *empilements* de rectangles de même largeur appartenant à différents profils. Cela revient à envisager toutes les cases en écart positif (ou négatif) d'une ligne du tableau d'origine. Par exemple on a vu qu'en examinant l'empilement *Pères agriculteurs* on a repéré que les agriculteurs avaient eu des enfants plutôt agriculteurs ou ouvriers. On peut vérifier sur le graphique (et le montrer formellement) que pour un empilement donné, la somme des surfaces au-dessus de la moyenne est égal à la somme des surfaces en-dessous ; de même pour un profil donné, la somme des surfaces en écart positif est égal à la somme des surfaces en écart négatif. Tout cela vient du fait que dans un tableau quelconque, la somme

(algébrique²⁹) des écarts positifs à l'indépendance est nulle pour une ligne donnée (empilements) ou pour une colonne donnée (profils)

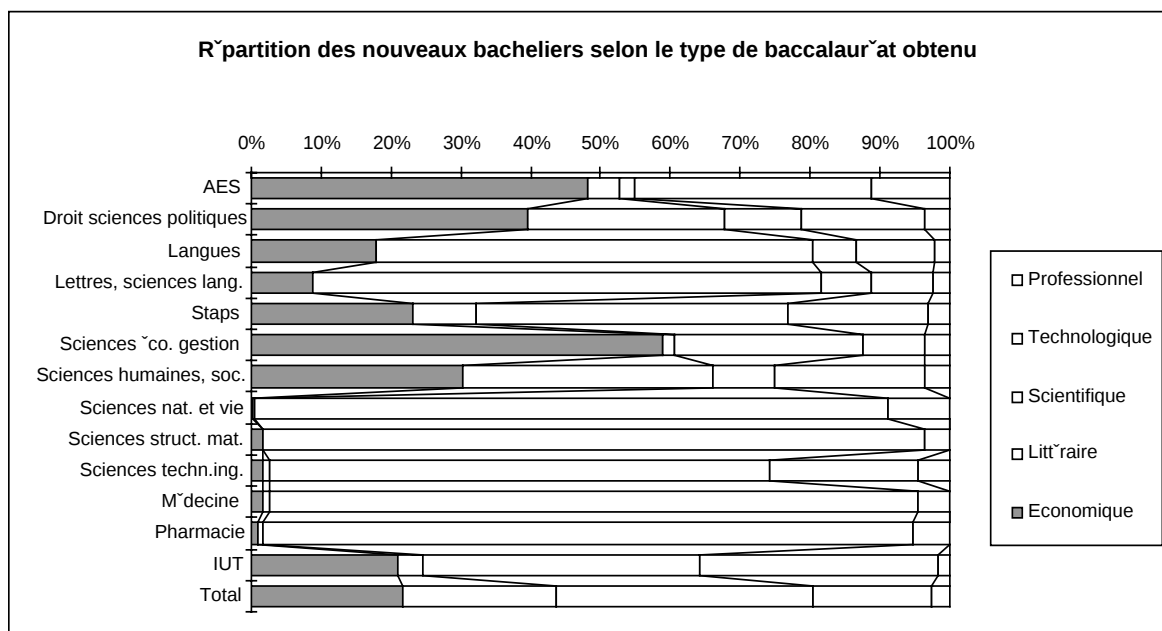
Nouveaux bacheliers

Examinons, dans le tableau NBAC0, la population des 257118 nouveaux bacheliers de 1997 selon le type de baccalauréat obtenu et selon l'orientation qu'ils ont prise à l'université³⁰.

Afin de comparer les modes de représentation, on trouvera dans le graphique NBAC1 la représentation traditionnelle des documents statistiques telle qu'elle était dans le document d'origine. La lecture d'un tel graphique est rendue difficile par le fait que les types de baccalauréats, sauf pour la série économique à gauche et le bac professionnel à droite, ne sont pas à l'aplomb. Un tel type de graphique permet de répondre à des questions élémentaires comme "quel est la proportion de série Economique et social en AES", mais il ne donne que très partiellement une vision d'ensemble.

²⁹ c'est à dire en tenant compte des signes *plus* ou *moins*.

³⁰ Sources : *Note d'information* 98.09 du Ministère de l'Education nationale



NBAC1

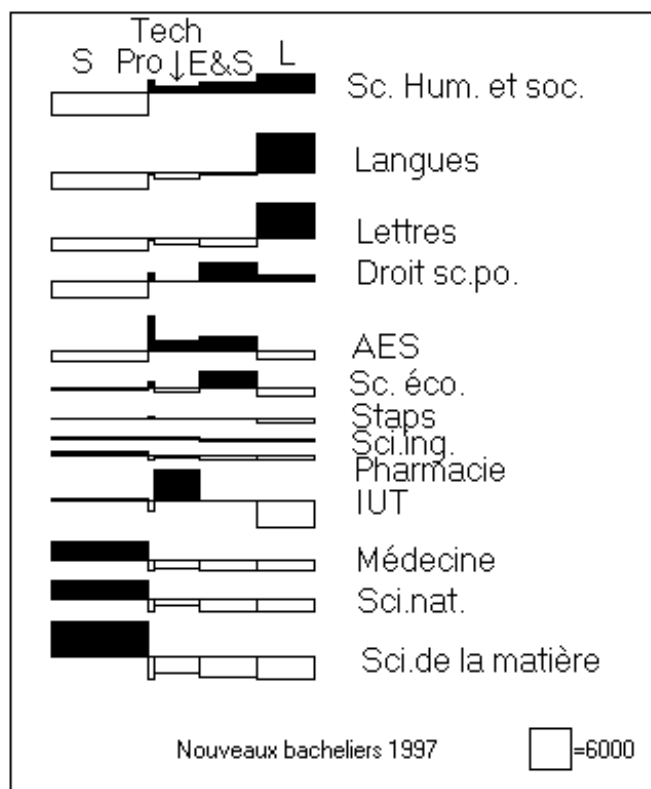
Pour pouvoir lire l'ensemble des données on utilise la représentation en surface du graphique NBAC2.

Nouveaux bacheliers 1997 : type de bac et discipline à l'université

Disciplines														
Type de bac	Sc.hum et soc.	Langues	Lettres	Droit, sciences po	AES	Sc.éco. gestion	STAPS	Sc.tech ingénieur	Pharmacie	IUT	Médecine	Sc. de la nature	Sc. de la matière	Total
S	3397	1699	1377	3302	311	3377	4141	1981	3945	16198	13526	14368	27175	94797
	3,6	1,8	1,5	3,5	0,3	3,6	4,4	2,1	4,2	17,1	14,3	15,2	28,7	100
Pro	1359	583	448	1083	1506	450	298	124	0	730	0	0	0	6581
	20,7	8,9	6,8	16,5	22,9	6,8	4,5	1,9	0,0	11,1	0,0	0,0	0,0	100
Tech	8152	3058	1721	5359	4516	1126	1855	594	228	13864	650	1417	1025	43565
	18,7	7,0	4,0	12,3	10,4	2,6	4,3	1,4	0,5	31,8	1,5	3,3	2,4	100
Eco & soc	11550	4854	1722	12016	6452	7429	2153	49	38	8610	260	28	513	55674
	20,7	8,7	3,1	21,6	11,6	13,3	3,9	0,1	0,1	15,5	0,5	0,1	0,9	100
L	13588	16989	14013	8552	597	225	828	25	38	1459	130	57	0	56501
	24,0	30,1	24,8	15,1	1,1	0,4	1,5	0,0	0,1	2,6	0,2	0,1	0,0	100
Total	38046	27183	19281	30312	13382	12607	9275	2773	4249	40861	14566	15870	28713	257118
	14,8	10,6	7,5	11,8	5,2	4,9	3,6	1,1	1,7	15,9	5,7	6,2	11,2	100

Source : Note d'Information 98.09
du MENRT

NBAC0



NBAC2

Le graphique NBAC2 permet de repérer :

1) des ressemblances entre les disciplines universitaires : on voit que les sciences de la matière, les sciences naturelles et la médecine recrutent plutôt des bacheliers scientifiques et ont un déficit pour les autres types de bac. La ressemblance visuelle des profils manifeste à la fois les attractions et les répulsions semblables mais aussi leur importance numérique équivalente. Ce qui les différencie de la pharmacie et des sciences de l'ingénieur où la faible importance numérique dissuade de comparaisons avec de trop faibles effectifs. Ces profils de faible importance numérique, qui semblent assez plats, manifestent par ce biais leur peu d'écart à l'indépendance. Cependant il

faut faire attention au fait que ces faibles écarts, s'ils peuvent être dus à une faible importance numérique peuvent être dus aussi au fait qu'une discipline recrute assez indifféremment dans tous les types de bac. A cette fin on peut comparer les STAPS (Sciences et techniques des activités physiques et sportives) et sciences éco. qui ont des effectifs du même ordre (4 à 5%) mais où on voit un profil plat pour les STAPS et des différenciations marquées pour les sciences éco. Contrairement aux sciences éco. qui recrutent de manière privilégiée dans la série économique et sociale du bac, les STAPS recrutent dans tous les types de bac : ceci vient du fait que cette section, dont le développement est récent, attire des sportifs qui peuvent venir de toutes les formations, indépendamment des hiérarchies de niveau entre sections. En haut du graphique on repère les similitudes fortes entre Langues et Lettres

2) le graphique permet de repérer des *spécificités* : par exemple on dit souvent que les IUT ne répondent pas à leur vocation première qui est d'accueillir les séries technologiques du bac. On voit sur le graphique la spécificité des IUT qui, bien plus que la moyenne, accueillent ces types de bac, tout en accueillant aussi légèrement plus que la moyenne les bac scientifiques, attraction qui leur vaut leur réputation sélective et élitiste.

En plus des spécificités déjà notées des bacs scientifiques pour les sciences de la matière et de la vie, et celles des lettres pour les Lettres et Langues, on repère plutôt des cas particuliers :

- que si sciences éco. est comparable à AES (Filière administrative, économique et sociale), on note l'attraction entre sciences éco. et section ES commune aux deux, mais par contre la spécificité d'AES qui recrute aussi dans les bacs technologiques. On peut comparer également avec la situation du Droit, fort recruteur en ES mais aussi en Lettres. Enfin, la filière sciences humaines et sociale cumule les attractions avec tous les types de bac sauf les sciences : c'est souvent

par un choix négatif contre les mathématiques que se fait le choix de cette filière.

- les spécificités peuvent être repérés dans l'autre sens en termes d'*empilements*. On peut voir par exemple que les bacs pro. peu nombreux d'ailleurs (moins de 3%) sont en attraction forte avec les AES, puis avec sciences humaines et sociales et droit. Cette série est en opposition tant avec les Lettres et Langues qu'avec les disciplines de sciences exactes. La filière des bacs professionnels, dernière venue du système, scolarise en dehors de la filière noble actuelle (les sciences) et même en dehors de l'ancienne filière noble (les lettres) et donne un accès (limité) aux disciplines d'une nouvelle culture (sciences humaines et sociales).

6 Force de l'attraction

Le pourcentage de l'écart maximum

Sur le graphique *Religion et politique* précédent (RELPOL6), on voit que les attractions entre « gauche » et « tradition catholique » ou « sans religion » correspondent à des effectifs d'individus en écart à l'indépendance assez proches. Peut-on aller plus loin et estimer la *force* de l'attraction en lui donnant une valeur numérique ?

Pour répondre à cette question, il faut se donner des bornes et replacer numériquement l'observation entre le cas d'une attraction nulle qui aura la valeur zéro et une liaison maximum qui aura par convention la valeur 100.

Prenons le cas de l'attraction « gauche », « sans religion » : isolons cette case du tableau en la donnant simplement avec ses marges et le total :

	Sans religion		Total
Gauche	149		448
.....			
Total	220		955

Attraction nulle : la réponse, dans le cas de l'absence d'attraction, est simple car on observerait pour cette case un effectif le plus proche possible de l'effectif théorique correspondant à l'indépendance (produit des marges divisé par le total, cf. plus haut).

$$\text{Effectif théorique} = 220 \times 448 / 955 = 103,2$$

L'observé le plus proche de l'indépendance serait donc de 103, or on observe un effectif de 149, donc on n'est pas proche de l'indépendance.

Attraction maximum : mais est-on proche de la liaison maximum ? Dans le cas présent, la liaison serait maximum si :

- tous les gens de gauche étaient sans religion
- ou si tous les sans religion étaient de gauche

La première option est incompatible avec les marges des données observées puisqu'elle supposerait que les 448 de gauche soient sans religion et il n'y a dans notre enquête que 220 sans religion. Par contre la deuxième est possible puisque les 220 sans religion peuvent être dans les 448 de gauche. D'une manière plus formelle, on voit qu'il suffit de mettre comme effectif dans le cas du maximum *la plus petite* des deux marges.

Si on avait un effectif de 220, comme l'effectif théorique correspondant à l'indépendance est de 103,2, on aurait un écart à l'indépendance de $220 - 103,2 = 116,8$ or l'écart à l'indépendance observé est de $149 - 103,2 = 45,8$

Ce que représente cet écart observé de 45,8 par rapport au maximum de 116,8 nous est donné par sa proportion de $45,8 / 116,8 = 0,392$ c'est à dire un pourcentage de 39,2%. Ce pourcentage de l'écart observé par rapport à l'écart maximum sera appelé dans la suite *Pourcentage de l'Ecart Maximum* ou PEM, ici égal à 39,2%.

Résumons les calculs pour l'autre case étudiée : « gauche » - « tradition catholique » :

	Tradition		Total
Gauche	170		448
.....			

Total	296	955
-------	-----	-----

- Effectif observé : 170
- Effectif théorique : $296 \times 448 / 955 = 138,9$
- Ecart observé : $170 - 138,9 = 31,1$
- Effectif maximum : 296
- Ecart maximum : $296 - 138,9 = 157,1$
- PEM : $31,1 / 157,1 \times 100 = 19,8\%$

Dans le cas des sans religion, le PEM était de 39,2%, ici il n'est plus que de 19,8% soit une valeur beaucoup plus faible, que le graphique ne nous permettait pas de repérer. Calculons les PEM pour toutes les attractions positives du tableau :

	Pratiquant	Non prat	Tradition	Sans religion
Gauche			19,8%	39,2%
Centre	3,8%	9,2%		
Droite	26,6%	17,7%		

On voit sur ce tableau des PEM que les plus fortes liaisons correspondent aux extrêmes ; que les sans religion sont plus liés à la gauche que les catholiques de tradition ; que les pratiquants sont plus liés à la droite que les non pratiquants ; quand au centre, il correspond

à des liaisons très faibles. Cependant, il n'y a pas symétrie, les sans religion sont plus liés à la gauche que les pratiquants à la droite ce qui manifeste que la gauche est davantage une culture de rupture, d'opposition (donc plus radicale) que la droite qui est une culture de *statu quo*.

Les valeurs du PEM varient de 3,8% à 39,2%. L'expérience³¹ montre que :

- en dessous de 5% les PEM correspondent à des attractions en général négligeables,
- que 10% constitue un seuil : au-dessus, les PEM sont en général intéressants, entre 5% et 10%, les liaisons peuvent l'être ou non,
- que des liaisons supérieures à 50% sont exceptionnelles et manifestent souvent qu'on a affaire à deux indicateurs d'un même phénomène. Si, dans un tableau 2×2 on croise une question avec elle-même, les PEM de la diagonale où se trouve les effectifs, sont évidemment égaux à 100%.

7 Application du PEM

Le PEM est un indicateur simple à calculer, facile à interpréter puisque c'est un pourcentage, mais son utilité se révèle surtout quand il est utilisé d'une manière systématique dans un logiciel de dépouillement d'enquête : il permet en effet d'explorer automatiquement un grand nombre de tris croisés et de sélectionner,

³¹ L'expérience des utilisateurs est indispensable pour chaque usage d'un pourcentage : ainsi celui qui grimpe une côte à vélo sait que jusqu'à 5% une pente ne pose pas de problèmes, que de 5 à 9% elle est notable, que de 9 à 13% elle est redoutable et qu'au dessus de 13%, il faut être très bien entraîné. Par contre quand on parle de croissance de l'économie d'un pays, 3% est bien mais 4% (et plus), est exceptionnel. Chaque usage de pourcentages a sa plage d'utilisations spécifiques.

pour chacune des modalités de réponse de l'enquête, les autres modalités qui lui sont spécialement liées au sens du PEM.

Prenons à titre d'exemple un extrait de l'enquête REL86 déjà citée : nous prenons la sous-population des 1432 baptisés catholiques pour explorer les attractions entre les questions suivantes :

Vous-même, vous considérez-vous comme : un croyant convaincu (434), un croyant par tradition (386), un croyant incertain (194), un sceptique (217), un incroyant (179), non-réponse (22).

Cette question forme une échelle en cinq points qui indique bien le niveau de croyance ou des difficultés de la croyance. On prend ensuite une question située au centre des croyances catholiques, la croyance en la résurrection du Christ à laquelle on demandait si les gens y croyaient (oui : 1023, non : 290, non-réponse : 119)

On utilise ensuite une question touchant le désir d'être enterré religieusement ou non, dont on sait que même pour beaucoup de gens nés catholiques mais ayant pris leurs distances vis-à-vis de l'Eglise, ce geste reste important. A la question "souhaitez-vous être enterré religieusement ?", 1056 répondent oui, 221 non (155 non-réponses).

Sur le plan de la morale catholique, on teste le point de vue suivant : "pensez-vous que l'on puisse se considérer comme catholique et accepter le principe des relations sexuelles avant le mariage ?". 1023 pensent que oui, cela est possible, 290 non (119 non-réponses).

Enfin, pour éclairer ces réponses par des modalités de statut social, on utilise aussi les questions suivantes : sexe (masculin 679, féminin 753), âge (en 5 tranches d'âge) et une question portant sur le parti politique dont on se sent le plus proche où l'on a recodé d'une part les proximités de gauche (567), d'autre part celles de droite (506), ce qui laisse 359 individus qui refusent de se classer politiquement.

On a donc en tout 7 questions : si on envisageait un croisement de ces questions deux à deux, il y aurait 21 tableaux différents dont il

faudrait examiner les attractions intéressantes. Plutôt que de tenter cette exploration, on fait calculer ces tableaux d'une manière logicielle, on en fait calculer tous les PEM et, pour chaque modalité, on donne la liste des autres modalités de l'enquête en attraction avec cette modalité, par ordre décroissant de PEM.

Question Croyance, modalité croyant convaincu

78,9%	Souhaite être enterré religieusement
76,8%	Croit en la résurrection du Christ
30,2%	Est politiquement proche de la droite
26,8%	On ne peut pas être catholique et accepter les relation sexuelles avant mariage
20,3%	Sexe féminin
17,8%	60 ans et plus
10,7%	45-59 ans

Ces PEM sont issus de 6 tableaux croisés différents : la croyance, croisée avec toutes les autres questions. Par exemple soit le tableau croisant croyance et enterrement qui permet de calculer le premier PEM.

Croyance	Souhait d'être enterré religieusement			
	Oui	Non	Non-rep	Total
Convaincu	410	10	14	434 (30,3)
	94,5	2,3	3,2	100
Tradition	343	15	28	386 (27,0)
	88,9	3,9	7,3	100
Incertain	153	20	21	194 (13,5)
	78,9	10,3	10,8	100
Sceptique	108	55	54	217 (15,2)
	49,8	25,3	24,9	100
Incroyant	29	120	30	179 (12,5)

	16,2	67,0	16,8	100	
Non-réponse	13	1	8	22	(1,5)
	59,1	4,5	36,4	100	
Total	1056	221	155	1432	(100)
	73,7	15,4	10,8	100	

On isole la ligne "convaincu", la colonne "oui" et les marges et on a l'extrait de tableau suivant :

	"Oui"	Total
Convaincu	410	434

Total	1056	1432
-------	------	------

- Effectif observé : 410
- Effectif théorique : $1056 \times 434 / 1432 = 320,0$
- Ecart observé : $410 - 320,0 = 90,0$
- Effectif maximum : 434
- Ecart maximum : $434 - 320,0 = 114,0$
- PEM : $90,0 / 114,0 \times 100 = 78,9\%$

Si nous examinons la modalité "convaincu" avec les sept modalités en attraction plus ou moins forte avec elle, on voit que les attractions ont été rangées par ordre de PEM décroissant, les premières étant les plus fortes comme l'attraction avec le souhait d'enterrement religieux très élevé avec un PEM de près de 80%.

On peut considérer l'ensemble des modalités associées à "convaincu" comme le *profil* de cette modalité : les modalités qui s'y trouvent "éclairent" la modalité de référence du fait qu'elles ont un lien privilégié avec elle. Il s'agit d'une explication à l'intérieur de l'univers des modalités de l'enquête mais il n'y a là aucun inconvénient. Il en est de même dans un dictionnaire où un mot est expliqué par d'autres mots de ce même dictionnaire, ceux qui sont en relation sémantique privilégiée avec lui.

Ici, on voit que le fait d'être "convaincu" est lié non seulement à l'enterrement à l'Eglise mais aussi d'une manière aussi forte avec une croyance centrale du dogme catholique, la résurrection du Christ, avec le fait d'être politiquement à droite, de ne pas accepter les relations sexuelles avant le mariage, d'être de sexe féminin et d'une des deux tranches d'âge les plus âgées de l'enquête.

Continuons la suite de l'exploration des profils de modalités par PEM décroissants :

Question Croyance, modalité croyant par tradition

- 57,6% Souhaite être enterré religieusement
- 14,7% On peut être catholique et accepter les relation sexuelles avant mariage
- 12,5% Non-réponse sur la croyance en la résurrection du Christ
- 8,9% Sexe masculin
- 7,7% Croit en la résurrection du Christ
- 4,7% 60 ans et plus

Cette fois s'il s'agit de personnes plutôt âgées, elles sont de sexe masculin : elles sont également attachées aux obsèques religieuses mais sur les questions de dogme soit elles hésitent, soit leur attraction est faible. Par contre, on voit apparaître l'acceptation des relations sexuelles avant le mariage. Sauf pour l'enterrement religieux, les attractions sont beaucoup moins marquées que dans la modalité "convaincue".

Question Croyance, modalité croyant incertain

- 20,6% On peut être catholique et accepter les relation sexuelles avant mariage

19,5% Souhaite être enterré religieusement

18,5% Sexe féminin

8,4% 25-34 ans

7,8 35-44 ans

Question Croyance, modalité croyant sceptique

41,0% Ne croit pas en la résurrection du Christ

38,7% On peut être catholique et accepter les relation sexuelles avant mariage

23,7% Se situe politiquement à gauche

23,2% Non-réponse à la question sur l'enterrement religieux

15,0% Sexe masculin

11,7% ne souhaite pas être enterré religieusement

10,3% Non-réponse à la question sur la croyance en la résurrection du Christ

9,3% 25-34 ans

7,4% 24 ans et moins

Question Croyance, modalité incroyant

80,2% Ne croit pas en la résurrection du Christ

61,0% Ne souhaite pas être enterré religieusement

37,1% Se situe politiquement à gauche

28,8% Sexe masculin

12,8% 25-34 ans

7,8% Non-réponse à la question sur l'enterrement religieux

6,7% 24 ans et moins

L'incroyance augmente avec un âge jeune, une orientation politique à gauche. On constate qu'on retrouve avec la modalité la plus extrême du côté de l'incroyance, des niveaux élevés de PEM sur le refus d'une croyance centrale comme la Résurrection où le refus d'être enterré religieusement. La force des convictions trouve son parallèle avec la force du PEM qui est toujours beaucoup plus fort quand il s'agit de lier des croyances et des opinions que quand on lie des opinions et des caractéristiques sociales.

On peut continuer ainsi à explorer l'ensemble des profils. Chaque PEM s'y trouve deux fois : par exemple si nous regardons les deux profils de gauche et de droite, on y retrouvera des liaisons déjà rencontrées plus haut (avec les modalités de croyance).

Question Proximité politique, modalité droite

- 43,5% Souhaite être enterré religieusement
- 36,2% Croit en la résurrection du Christ
- 30,2% Croyant convaincu
- 12,0% On ne peut pas être catholique et accepter les relations sexuelles avant mariage

Question Proximité politique, modalité gauche

- 37,1% Se considère incroyant
- 37,1% Ne souhaite pas être enterré religieusement
- 23,7% Se considère sceptique
- 22,8% On peut être catholique et accepter les relation sexuelles avant mariage
- 21,7% Ne croit pas en la résurrection du Christ
- 10,3% Non-réponse à la question sur l'enterrement religieux
- 8,1% Sexe masculin
- 9,3% 25-34 ans

Ces deux profils rejoignent ce que nous avons dit précédemment sur les rapports entre politique et religion : on remarquera que si les deux profils sont cohérents l'un avec la croyance, l'autre avec son refus, l'enterrement religieux est une question qui laisse sans réponse des gens de gauche (ce dont on a trouvé une illustration dans les dernières volontés d'un président de la République de gauche qui envisageait cette éventualité comme possible). L'enterrement religieux

est certes un rite religieux, et est refusé ou accepté comme tel, mais sa signification dépasse l'affiliation à une croyance religieuse³².

La technique des profils de PEM est la manière rapide d'explorer de nombreux tableaux croisés : en faisant ainsi on trouve beaucoup d'attractions dont on ne juge pas qu'elles apportent des informations nouvelles, mais on peut ainsi rapidement repérer des choses intéressantes.

Cette technique, mise au point pour étudier des liaisons entre modalités peut maintenant être étendue pour étudier des modalités entre questions dans leur ensemble.

8 Le PEM généralisé

On généralise la démarche du PEM en se demandant, non plus pour une case mais pour un tableau tout entier, ce que serait la liaison maximale et quelle serait dans ce cas la valeur de l'ensemble des écarts positifs à l'indépendance. Inversement on a facilement la situation correspondant à l'absence de liaison puisque c'est l'indépendance. Il restera à placer l'observation, synthétisée par la valeur de la somme des écarts positifs à l'indépendance, entre l'absence de liaison et la liaison maximum. Le PEM généralisé sera le rapport entre la somme des écarts positifs observés et la même somme dans le cas de la liaison maximum.

Reprenons le tableau croisant opinion politique et affiliation religieuse :

³² Sur cette question cf les travaux de Jean-Hugues Dechaux et par exemple *Le souvenir des morts : essai sur le lien de filiation*, Paris, Presses Universitaires de France, 1997.

	Pratiquant	Non prat	Tradition	Sans religion	Total
Gauche	34	95	170	149	448
Centre	51	113	93	52	309
Droite	61	85	33	19	198
Total	146	293	296	220	955

Par l'étude des PEM locaux (c'est à dire case par case), nous avons déjà observé que les attractions entre modalités se faisaient de la façon suivante:

	Pratiquant	Non prat	Tradition	Sans religion
Gauche			19,8%	39,2%
Centre	3,8%	9,2%		
Droite	26,6%	17,7%		

c'est à dire en suivant une diagonale du tableau (haut-droit à bas-gauche). Cette diagonale nous indique la forme générale de la liaison : aux extrémités, les liaisons sont fortes entre *gauche* et *sans religion* et entre *droite* et *pratiquant*. Si la liaison était à son maximum, c'est sur cette diagonale, ou à proximité d'elle, que se concentreraient les effectifs observés. Par exemple, en commençant par la case située en haut et à droite du tableau, si la liaison était à son maximum, on pourrait y mettre les 220 *sans religion*. Pour cette colonne, tous les effectifs seraient dans la ligne intitulée *gauche*, mais celle-ci ne serait pas épuisée pour autant puisqu'il en resterait $448 - 220 = 228$ à placer au plus près. Ils peuvent l'être dans la colonne *tradition* ce qui fait que la ligne *gauche* est entièrement placée sur la diagonale ou au plus près. Sont déjà placés les effectifs suivants :

	Pratiquant	Non prat	Tradition	Sans religion	Total
Gauche	0	0	228	220	448
Centre	-	-	-	0	309
Droite	-	-	-	0	198
Total	146	293	296	220	955

On alterne lignes et colonnes : après avoir terminé la colonne *sans religion*, puis la ligne *gauche*, on termine la colonne *tradition* en mettant le plus près possible de la diagonale du tableau l'effectif restant de $296 - 228 = 68$ que l'on met dans la ligne intitulée *centre*. Il reste $309 - 68 = 241$ du *centre* que l'on place près de la diagonale c'est à dire dans *non pratiquant*. Il reste $293 - 241 = 52$ *non pratiquant* que l'on place à *droite* et il reste enfin $198 - 52 = 146$ *droite* que l'on place dans *pratiquant*. Le processus est terminé avec le résultat final suivant, qui aurait été le même si on était parti du bas de la diagonale.

	Pratiquant	Non prat	Tradition	Sans religion	Total
Gauche	0	0	228	220	448
Centre	0	241	68	0	309
Droite	146	52	0	0	198
Total	146	293	296	220	955

Les marges n'ont pas changé par rapport au tableau de départ. Ce sont elles qui nous servent de cadre de référence, tant pour l'absence de liaison (effectifs théoriques correspondants à l'indépendance) que pour la liaison maximum (où l'on cherche à placer ces effectifs de marge au plus près de la diagonale de référence).

L'observation préalable des liaisons est indispensable pour savoir quelle est la diagonale de référence : de plus il est indispensable, pour que les diagonales soient bien identifiées, qu'il existe un ordre sur les lignes du tableau. Cette procédure du chargement d'une diagonale est toujours possible et conduit toujours au même résultat, une fois que la diagonale de référence a été choisie au vu de l'observation.

Le tableau d'indépendance est le suivant (c'est le même que celui du tableau d'origine puisqu'ils ont tous les deux les mêmes marges.)

	Pratiquant	Non prat	Tradition	Sans religion	Total
Gauche	68,5	137,4	138,9	103,2	448
Centre	47,2	94,8	95,8	71,2	309
Droite	30,3	60,7	61,4	45,6	198
Total	146	293	296	220	955

Les écarts à l'indépendance par rapport à l'observation sont :

	Pratiquant	Non prat	Tradition	Sans religion
Gauche	-34,5	-42,4	31,1	45,8
Centre	3,8	18,2	-2,8	-19,2
Droite	30,7	24,3	-28,4	-26,6

La somme des écarts positifs de ce tableau est de 153,9

Les écarts à l'indépendance par rapport à la liaison maximum sont :

	Pratiquant	Non prat	Tradition	Sans religion
Gauche	-68,5	-137,4	89,1	116,8
Centre	-47,2	146,2	-27,8	-71,2
Droite	115,7	-8,7	-61,4	-45,6

La somme des écarts positifs de ce tableau est de 467,9

Là proportion des écarts observés par rapport aux écarts dans le cas du maximum est de $153,9 / 467,9 = 0,329$ soit un PEM généralisé de 32,9%. On dispose ainsi d'un indicateur de liaison entre questions ordonnées (et on peut toujours les ordonner³³). Cet indicateur servira dans la suite pour sélectionner rapidement, dans une enquête, des questions qui sont plus ou moins liées entre elles.

³³ par exemple en utilisant l'ordre du premier facteur d'une analyse factorielle : c'est ce que fait le logiciel Trideux.

9 Application du PEM généralisé

Quittons les exemples à finalité pédagogique pour voir comment le PEM généralisé peut être utilisé dans une recherche en vraie grandeur : il s'agit d'une enquête³⁴ sur le choix par des parents de faire faire du latin à partir de la classe de 4^e. Les données utilisées sont celles d'une enquête portant sur une cohorte d'élèves suivis par les services statistiques du Ministère de l'éducation nationale depuis leur entrée en 6^e en 1989³⁵. Pour un certain nombre de ces élèves on possède un certain nombre de renseignements sur leur réussite scolaire mais aussi sur la manière dont ils organisent leur journée, leur semaine, leurs vacances et des renseignements sur leurs parents. On dispose ainsi de plus de 250 questions : pour sélectionner les questions qui sont le plus liées au choix de l'option latin en 4^e, on calcule le PEM généralisé entre cette question précise et l'ensemble de toutes les autres questions de l'enquête : on obtient là aussi un *profil* de questions liées par PEM décroissant. Soit par exemple des extraits de la liste par ordre d'importance décroissant du PEM global avec sa valeur, l'intitulé des questions et celui des modalités de réponse.

Question Choix ou non de l'option latin en 4^e

88% Durant les vacances d'été, t'est-il arrivé de suivre des sessions de rattrapage scolaire ? (oui/non)

82% Les parents sont-ils membres d'une association de parents d'élèves ? (oui/non)

78% Jusqu'à quel âge les parents souhaitent-ils que leur enfant poursuive ses études ? (16, 18, 20 et plus)

³⁴ Cf. Philippe Cibois, "Le choix de l'option latin au collège", *Education & Formations*, n°48-1996, p.39-51

³⁵ Cohorte appelée le PANEL89

63% Age (se trouve en avance, à l'âge normal, en retard)

63% Niveau en mathématiques (bon, moyen, très passable, insuffisant)

58% Les parents aident-ils habituellement leur enfant dans son travail scolaire ? (oui, parce qu'il rencontre des difficultés ; oui, bien qu'il n'ait pas de difficultés ; non, parce qu'il n'a pas besoin d'aide ; non, par manque de temps ; non pour une autre raison)

Cette procédure des profils de PEM généralisé permet de repérer les questions qui sont en rapport, elle ne permet pas de savoir dans quel sens se fait la liaison. Prenons par exemple les deux plus forts PEM et allons voir les tableaux croisés dont ils sont issus :

1) PEM de 88% entre le fait d'avoir choisi ou non l'option latin et le fait d'avoir ou non suivi des cours de rattrapage pendant les vacances :

		Option latin		
		non	oui	Total
Rattrapage pendant les vacances	oui	29	1	30
	non	1497	338	1835
Total		1526	339	1865

Sur cet exemple on voit que la liaison se fait presque parfaitement entre le fait de faire du latin et le fait de ne pas avoir suivi de rattrapage pendant les vacances. La liaison se fait entre le latin et la non-pratique : ceux qui choisissent l'option latin n'ont pas besoin d'aide car ce sont de bons élèves.

Dans l'exemple suivant, on croise le latin avec l'appartenance des parents à une association de parents.

		Option latin		
		non	oui	Total
Parents membre d'une association de parents d'élèves	oui	159	113	272
	non	1367	226	1593

Total	1526	339	1865
-------	------	-----	------

La situation est ici différente : le pourcentage de l'option latin est de 41,5% dans les familles qui sont membres d'une association, il n'est que de 14,2% pour les autres. Il y a une attraction entre le fait de faire du latin et le fait d'être membre d'une association. On voit que les parents de latinistes sont plus intégrés dans le système scolaire qu'ils connaissent mieux que les autres.

On verrait d'une manière similaire que ceux qui font l'option latin sont plutôt en avance, qu'ils sont bons en mathématiques, que leurs parents souhaitent leur faire faire des études longues et qu'ils ne sont pas aidés parce qu'ils n'en n'ont pas besoin. Mais ces résultats ne viennent pas directement du profil des PEM globaux qui ne servent qu'à sélectionner les questions pertinentes, mais du profil de PEM local (lié à une case d'un tableau, le premier PEM que nous ayons étudié) obtenu en travaillant sur l'ensemble des *modalités* cette fois engendré par l'ensemble des *questions* sélectionnées. Le PEM global constitue la première étape d'une recherche dans un dépouillement d'enquête, ensuite on passe au PEM local associé aux modalités.

Voici donc à titre complémentaire le profil de PEM *local* associé à la modalité : a choisi l'option latin en 4^e

Modalité "a choisi l'option latin en 4^e"

- 43% A l'heure ou en avance à l'entrée en élémentaire
- 39% Niveau en mathématiques le plus élevé
- 35% A l'heure à l'entrée en sixième
- 34% A été au théâtre depuis le début de l'année
- 30% Poursuite des études envisagée après 20 ans
- 28% A été au concert depuis le début de l'année
- 28 % diplôme mère : grandes écoles ou 2e cycle univ.
- 25 % a été à l'étranger à Pâques
- 24 % diplôme père : grandes écoles ou 2e cycle univ.

- 24 % suit des cours de musique
- 24 % parent délégué
- 23 % aime d'autres lectures que les lectures scolaires
- 23 % lecture loisir favori
- 21 % a lu plus de 10 livres depuis le début de l'année
- 21 % études utiles jusqu'à un diplôme d'enseignement supérieur
- 20 % parent membre association de parents d'élèves
- 19 % est inscrit à un conservatoire de musique
- 19% chef de famille cadre ou prof. intellectuelles supérieures
- 18 % est très intéressé par les langues vivantes
- 17 % fait de l'allemand comme première langue vivante
- 17 % est très intéressé par l'histoire
- 16 % le collège a été choisi pour sa bonne réputation
- 15 % n'oublie jamais ses affaires
- 15 % est très intéressé par le français
- 14 % a été au cinéma depuis le début de l'année
- 14 % est parti en février avec ses parents
- 13 % l'enfant vit avec son père et sa mère
- 13 % est aidé bien qu'il n'ait pas de difficultés
- 13 % est dans l'enseignement privé
- 13 % dispose d'une chaîne haute fidélité
- 13 % très intéressé par la géographie
- 13% chef de famille professeur
- 12 % le mercredi va à la bibliothèque
- 10 % trouve que lire c'est amusant
- 10 % est inscrit en bibliothèque
- 10 % a été à une exposition depuis le début de l'année

Si l'on fait la synthèse de toutes ces informations on peut dire que choisissent plutôt l'option latin en 4^e des élèves qui sont d'abord de bons élèves tant en mathématiques qu'en français. Dans le domaine littéraire, ils lisent beaucoup, y trouvent du plaisir et font d'autres activités culturelles en dehors de l'école (théâtre, musique en

particulier). Leurs parents, de niveau social élevé ont pour eux de grandes ambitions. L'enfant présente une "bonne volonté scolaire" (il n'oublie jamais ses affaires par exemple) qui le rend tout disposé à donner suite aux conseils de lecture de ses enseignants. Les parents, qui prennent à cœur leur "métier de parents" en s'investissant dans les associations ou en ayant fait un choix spécifique pour le niveau du collège (ou ont utilisé à cette fin l'enseignement privé). En conclusion on voit que cette option latin est liée à l'excellence scolaire d'élèves dont les parents insistent sur une culture littéraire (français et latin à l'école) et artistique (en dehors de l'école, mais dans la même logique).

10 Résumé et conclusion

Dans des données recueillies lors d'une enquête on a des *modalités* de réponses à des *questions* (appelées quelquefois aussi *variables*). À l'intersection d'une ligne et d'une colonne le nombre d'individus est dit *l'effectif observé*.

Une *attraction* est une association, plus que la moyenne entre deux modalités de réponse à deux questions. Une *répulsion* (ou *déficit*) est une association *moins* que la moyenne. Attractions et répulsions sont repérées par un écart à la moyenne. L'*écart à la moyenne* est la différence entre le pourcentage d'une ligne et le pourcentage toutes lignes confondues. Attractions, répulsions, écarts à la moyenne sont le reflet d'*écarts à l'indépendance*. Un *effectif correspondant à l'indépendance*, dit *effectif théorique*, correspond à une situation d'absence de lien, il est obtenu par produit des marges divisé par le total. L'*écart à l'indépendance* est la différence entre l'effectif observé et l'effectif théorique : il peut être visualisé sous forme graphique.

Le *pourcentage de l'écart maximum* (PEM) mesure, sous forme de pourcentage, la part que représente l'effectif en écart à l'indépendance avec l'effectif correspondant si la liaison était à son maximum. Le

PEM peut être soit *local* (à l'échelon d'une case d'un tableau) soit *global* au niveau de l'ensemble du tableau.

Avec ces outils on a les instruments pour explorer des données d'enquête où l'on a de nombreuses questions : dans une première étape on utilise le PEM global pour associer à chaque question les autres questions de l'enquête qui lui sont spécialement liées. Ensuite, pour des modalités particulières, on peut utiliser des profils de PEM locaux de modalités. On peut également visualiser des tableaux intéressants. Cependant, quand les effectifs sont faibles il est raisonnable de se demander si les liaisons observées ne seraient pas dues au hasard : c'est ce que nous allons étudier plus particulièrement.

Chapitre 2

Hasard et signification

1 *Les effets du hasard*

Savez-vous ce qui s'est passé d'intéressant à New-York en 1966 ? Probablement non, mais si je vous parle d'une panne de courant, il est probable que cela vous évoquera le fait que neuf mois après cette panne, il y a eu une augmentation des naissances. Depuis d'ailleurs, le fait s'est semble-t-il répété car à chaque panne de courant analogue, on prédit le même effet, voire même on l'annonce comme réalisé.

Or il n'en est rien : il s'agit d'une pure rumeur basée comme toutes les rumeurs sur une *attente* et sur un *fait déclencheur*. L'attente est le désir de rupture des interdits que pourrait procurer l'obscurité, comme naïvement un sociologue osa le déclarer au *New York Times*³⁶ : "La lumière s'est éteinte, et les gens ont été abandonnés à des interactions mutuelles". Quant à l'élément déclencheur ce fut celui-ci, toujours d'après le *New York Times* : "Une augmentation brutale du nombre des naissances est signalée par plusieurs grands hôpitaux new-yorkais, neuf mois après la grande panne d'électricité de 1965. A l'hôpital du Mont-Sinaï, la moyenne journalière est de 11 naissances, il y en eut 28 le 8 aout. C'est le record de cet hôpital ; le maximum enregistré antérieurement n'était que de 18. A l'hôpital Bellevue, 29 bébés sont nés le 9 aout contre seulement 11 une semaine auparavant, la moyenne n'étant que de 20. L'hôpital Columbia-Presbyterian dont la moyenne est de 11 naissances en a eu 15 le 8 aout. St-Vincent pour une

³⁶ les éléments de base de ce dossier sont issus de : Gérard Calot, "Les effets de la grande panne de courant à New York (1965)", *Economie et Statistique*, 1970, n°18.

moyenne de 7, en a eu 10 ; Brookdale : 10 en moyenne, 13 le 8 aout ; Coney Island : 5 en moyenne, 8 le 8 aout.". Pourtant, si l'on examine le nombre des naissances de l'agglomération de New York pendant les 6 semaines de juillet et aout 96 où auraient dû se passer les naissances liées à la panne, on ne trouve aucune variation par rapport aux années antérieures.

Classons ces mêmes données en considérant la moyenne des naissances par jour, l'observation du jour, l'écart par rapport à la moyenne et ce que représente cet écart par rapport à la moyenne (en pourcentage) :

	Moyenne	Observation	Ecart	Ecart/moyenne
Mont-Sinaï	11	28	+17	154%
Bellevue	20	29	+9	45%
Columbia Presb.	11	15	+4	36%
St-Vincent	7	10	+3	43%
Brookdale	10	13	+3	30%
Coney Island	5	8	+3	60%

Les écarts des 4 dernières cliniques qui sont de 3 ou 4 relèvent du quotidien, il n'y a là que des variations très ordinaires que l'on rencontre dans toute clinique. Pour la clinique de Bellevue, l'auteur de l'article a éprouvé le besoin d'amplifier le phénomène en mettant en rapport l'observation du 9 aout (et non du 8 comme pour les autres, ce qui laisse penser que celle du 8 était moins spectaculaire) avec celle, anormalement basse de la semaine précédente de 11. Sur une moyenne de 20, un écart de 9 ne représente que 45% ce qui classe cette clinique dans la moyenne des autres.

Seul un cas sort de l'ordinaire : il s'agit de la clinique du Mont-Sinaï où l'écart observé représente *une fois et demie* la moyenne alors que les autres écarts tournent aux alentours de *la moitié* de la moyenne. On peut faire l'hypothèse que tout est parti de cette clinique : les médecins sont submergés, les sages-femmes aux abois, les

patientes attendent dans des lits de fortune installés dans les couloirs. Pourquoi ? Serait-ce la pleine lune, source bien connue (bien que mythique³⁷) d'augmentation des naissances ? Non, celle-ci n'aura lieu que dans 4 jours, le 12 août. Alors un interne a un trait de génie : c'est la faute de la panne d'il y a neuf mois. Pour s'en assurer, il fait par téléphone le tour de ses collègues et ne retient évidemment que les écarts positifs à la moyenne, chiffres qui ne sont pas d'ailleurs mirifiques mais qui suffisent à faire la matière d'un coup de téléphone aux journaux qui sautent sur l'occasion.

Il a donc suffi d'un seul cas sortant réellement de l'ordinaire pour que se crée une légende devenue depuis une rumeur puisqu'elle se répète de panne en panne et fait l'objet d'une croyance bien établie. Ceci manifeste clairement que face à un événement qui sort beaucoup de l'ordinaire, l'on ne croie pas au hasard : on ne peut pas admettre qu'un événement exceptionnel arrive sans raison, simplement du fait que, *par hasard*, la clinique du Mont-Sinaï ait reçu une fois et demie de patientes de plus que la moyenne. Le statisticien peut se lamenter de ce fait³⁸ mais le sociologue doit plutôt en prendre acte : le très exceptionnel doit (socialement) s'expliquer autrement que par le hasard. Si je lance 10 pièces de monnaie sur la table et qu'elles retombent toutes sur *face*, vous me demanderez mon truc alors qu'il n'est pas impossible que cela arrive par hasard (avec une probabilité d'environ une chance sur mille). Quand la probabilité de ce qui arrive réellement devient trop faible, le hasard devient inacceptable et on

³⁷ Cf. pour le cas français : Frédéric Saudron, *Les naissances de la pleine lune*, Paris, L'Harmattan, 1998

³⁸ c'est ce que fait Gérard Calot dans l'article cité : "l'épisode nous paraît digne d'intérêt parce qu'il montre combien est sujet à caution le témoignage que tout individu pris isolément peut apporter sur la base d'une information nécessairement partielle (...) l'attention de chacun est naturellement attirée par l'exceptionnel, voire le sensationnel, et se détourne du banal".(p.47)

cherche une autre explication : c'est bien sur cette pratique que s'appuie l'utilisation des tests statistiques que nous allons étudier mais pour laquelle nous allons nous donner d'abord un instrument de mesure, l'écart pondéré.

2 L'écart pondéré

Imaginons une population de 1000 étudiants classés selon leur série du bac et selon leur destination l'année suivante : université, classes préparatoires aux grandes écoles et autres orientations (IUT et autres formations à finalités professionnelles).

Série	Université	Classes prep.	Prof.	Total
Littéraire	130	20	50	200
Eco.et soc.	200	20	80	300
Scientifique	100	50	50	200
Tech.et pro.	70	10	220	300

Total	500	100	400	1000

Dans ce tableau, isolons deux cases : "littéraires allant à l'université" et "scientifiques allant dans une classe préparatoire", et calculons pour chaque cas l'écart à l'indépendance :

	Littéraires Université	Scientifiques Classes préparatoires
Observé	130	50
Théorique	$500 \times 200 / 1000 = 100$	$100 \times 200 / 1000 = 20$
Ecart	$130 - 100 = 30$	$50 - 20 = 30$

L'écart pour les deux cases est bien le même, égal à trente individus, mais ce même écart à l'indépendance a-t-il la même signification dans les deux cas ? On peut considérer qu'il n'en est pas ainsi car pour les littéraires allant à l'université, l'effectif théorique calculé à partir des marges prend en compte l'importance de

l'université (la moitié des destinations). L'effectif théorique de ce fait est aussi un effectif "attendu" : c'est l'effectif qu'il y aurait en cas d'indépendance, effectif qui ne dépend que des marges. Puisque la moitié va à l'université, s'il y a indépendance, cette proportion s'applique aux 200 de la série littéraire soit 100 attendus.

On a donc un écart de 30 par rapport à une attente de 100 : l'écart par rapport à ce qui est attendu ne représente qu'une proportion de $30 / 100 = 0,3$ soit environ un tiers de l'effectif théorique. On attendait un certain effectif, on en observe un tiers en plus, l'évènement est de peu d'importance.

Par contre pour les scientifiques allant dans une classe préparatoire, comme cette destination ne représente que 10% des destinations, l'effectif théorique attendu ne représente que 10% des 200 de série scientifique soit 20. Or l'écart en plus est de 30 individus ce qui représente ici un rapport de $30 / 20 = 1,5$. L'écart est de une fois et demie l'effectif théorique. On attendait un certain effectif, il en arrive cette fois une fois et demie en plus, ce qui est tout à fait notable.

Dans le premier cas, l'écart de trente individus n'apporte pas beaucoup de surprise tandis que le même écart est très surprenant dans le deuxième cas : l'*information* apportée par le même écart n'est pas la même dans les deux cas. Pour manifester cette différence d'information apportée par les deux écarts, on construit un indicateur nouveau qui dérive de l'écart précédent par une opération de pondération, qui fasse en sorte que soit intégré dans le résultat numérique l'effet de surprise (dans le deuxième cas), ou son absence (dans le premier). A cette fin on se sert précisément du rapport écart/théorique (0,3 dans le premier cas, 1,5 pour le deuxième) comme coefficient de pondération pour diminuer ou amplifier l'écart, donc pour le pondérer.

Littéraire -
Université

Scientifique-
Classes préparatoires

Rapport écart/théorique	0,3	1,5
Ecart pondéré	$30 \times 0,3 = 9$	$30 \times 1,5 = 45$

Cet écart pondéré, produit d'un effectif (l'écart) par un coefficient de pondération (rapport de deux effectifs donc sans unité de compte) est un écart pondéré dont l'unité de compte peut être pensée comme un individu. Dans le premier cas, l'écart pondéré est de 9 individus, dans le deuxième il est de 45. Cet indicateur nous signale que dans le premier cas, l'écart brut de 30 individus n'apporte que peu de surprise (il est donc pondéré dans le sens de la diminution) tandis que dans le deuxième cas il est pondéré dans le sens de l'augmentation. Le fait d'aller dans les grandes écoles est un phénomène plutôt rare : les bacheliers venant de la série scientifique qui y vont, s'y retrouvent d'une manière tout à fait particulière, surprenante par son importance par rapport à l'effectif attendu et c'est d'ailleurs cette sélection par le petit nombre qui en fait tout l'intérêt dans le jeu social.

Cet écart pondéré est appelé le khi-deux d'une case. Le créateur de cette indice, Karl Pearson a employé pour le désigner une lettre grecque³⁹ mise au carré χ^2 , car cet indice est souvent présenté sous la forme suivante où un carré intervient :

$$(\text{observé} - \text{théorique})^2 / \text{théorique}$$

or "observé - théorique" correspond à l'écart à l'indépendance et le khi-deux peut s'écrire :

$$\text{écart}^2 / \text{théorique}$$

ou encore :

$$\text{écart} \times (\text{écart} / \text{théorique})$$

c'est à dire l'écart brut à l'indépendance pondéré par le rapport écart / théorique.

³⁹ les lettres grecques sont souvent utilisées par les mathématiciens pour désigner une valeur inconnue, dont on n'a qu'un "reflet", l'observation, en référence au mythe de Platon, où les hommes sont dans une caverne et où ce qu'ils voient ne sont que les reflets de la réalité idéale.

Le khi-deux d'une case est la même chose que l'écart pondéré : on peut ainsi considérer le khi-deux de toutes les cases d'un tableau comme la somme des écarts pondérés, la somme des "informations de surprise" faibles ou fortes apportées par les différents écarts du tableau. Un écart apporte toujours une certaine information, mais pour l'indice khi-deux, elle est pondérée par son rapport au théorique qui l'amplifie ou la réduit selon le cas. On passe de l'écart brut, ordinaire, à l'information apportée par cet écart, notée en nombre pondéré d'individus en écart à l'indépendance, avec donc une unité de compte homogène à un écart (et comparable à lui de ce fait).

3 Exemple d'utilisation : IUT

IUT : sexe - réussite

Pour utiliser un exemple réel soit le tableau qui croise, pour une population d'étudiants d'IUT⁴⁰, un indicateur de la réussite passée dans le secondaire (pas de redoublement ou un seul, opposé à deux redoublements ou plus) avec le sexe du répondant :

	0 ou 1 redoublement	2 redoublements ou +	Total
Masculin	1083	318	1401
Féminin	325	98	423
Total	1408	416	1824

En considérant le tableau d'indépendance, on s'aperçoit qu'il est très proche de l'observé:

	0 ou 1	2 redoublements	Total
--	--------	-----------------	-------

⁴⁰ d'après Philippe Cibois et Janina Lagneau, *Les étudiants dans l'enseignement supérieur court*, Paris, OCDE, 1976, p.110.

	redoublement	ou +	
Masculin	1081,5	319,5	1401
Féminin	326,5	96,5	423
Total	1408	416	1824

l'écart pour chacune des cases est en valeur absolue de 1,526. On est très proche de l'indépendance : d'ailleurs le PEM du tableau n'est que de 0,5%⁴¹ : calculons le khi-deux de chaque case en pondérant l'écart de 1,526 par le rapport *écart / effectif théorique* de chaque case soit : $1,526 \times 1,526 / 1081,5 = 0,002$ pour la première case :

	0 ou 1 redoublement	2 redoublements ou +	Total
Masculin	0,002	0,007	0,009
Féminin	0,007	0,024	0,031
Total	0,009	0,031	0,041

Comme les écarts sont minimes par rapport aux effectifs théoriques, la pondération des écarts a joué dans le sens de la diminution : les contributions de chaque case au khi-deux sont très faibles et le khi-deux total est de 0,041 soit largement moins qu'un individu. Quelque soit l'indicateur considéré : l'écart à l'indépendance proche de un individu, le PEM inférieur à 1%, le khi-deux inférieur à 1 individu ; tout nous signale que nous sommes très proche de l'indépendance et notre conclusion sera, sans que nous ayons ici besoin d'aller plus loin, qu'il y a indépendance entre le sexe et la réussite passée dans le secondaire. Dans notre population d'étudiants en IUT, la réussite dans le secondaire a été la même pour les garçons et les filles.

⁴¹ on peut parler de PEM du tableau, sans préciser si c'est un PEM local ou global puisqu'il s'agit d'un tableau 2 x 2 où ces deux PEM sont identiques.

Cependant nous observons des écarts par rapport à cette hypothèse, mais ils sont tellement faibles qu'ils ne remettent pas en cause l'hypothèse d'indépendance. Un écart pondéré inférieur à un individu peut raisonnablement être tenu pour négligeable. L'observation est trop proche de l'indépendance pour que nous ne prenions pas cette hypothèse comme guide d'interprétation : les deux sexes ont eu la même réussite dans le secondaire. Ce ne sera pas le cas avec la même population prise selon un autre point de vue.

IUT : origine sociale - réussite

Considérons maintenant en ligne non plus le sexe mais l'origine sociale (inférieure / supérieure) avec le même critère de réussite dans le secondaire en colonne :

	0 ou 1 redoublement	2 redoublements ou +	Total
Supérieure	697	265	962
Inférieure	711	151	862

Total	1408	416	1824

En considérant le tableau d'indépendance, on s'aperçoit maintenant qu'il est beaucoup moins proche de l'observé que dans le cas précédent :

	0 ou 1 redoublement	2 redoublements ou +	Total
Supérieure	742,6	219,4	1401
Inférieure	665,4	196,6	423

Total	1408	416	1824

L'écart pour chacune des cases est en valeur absolue de 45,6 et le PEM de 26,0% : nous sommes dans un cas très différent du précédent,

ce qui nous invite à regarder où se situent les attractions et les répulsions :

	0 ou 1 redoublement	2 redoublements ou +
Supérieure	-45,6	45,6
Inférieure	45,6	-45,6

On constate une attraction entre origine sociale *inférieure* et réussite scolaire dans l'enseignement secondaire attestée par le fait de ne pas avoir pris de retard (ou un an seulement) : comme ce résultat semble aller à l'encontre de ce qu'on sait sur la réussite scolaire pendant le secondaire, nous devons nous demander si nous ne sommes pas en présence d'un tableau peu fiable où des fluctuations d'échantillonnage de l'enquête nous interdiraient de donner trop d'importance aux résultats. Devons-nous remettre en cause les données ou notre interprétation ? Pour répondre à cette question préalable, regardons les écarts, qu'ils soient bruts ou pondérés.

L'écart brut est égal à 45,6 en valeur absolue, soit 91,2 pour la somme des écarts positifs : avoir dans un tableau une centaine d'individus en écart positif à l'indépendance n'est pas rien et ce simple fait nous incite à considérer le tableau comme pertinent, comme apportant une information, même si elle n'est pas immédiatement intelligible ; à considérer comme peu probable que nous soyons dans la même situation que dans le cas précédent avec la variable sexe et où il y avait indépendance. Cela semble raisonnablement peu probable mais comment sortir du registre de l'impression, du sentiment raisonnable, de l'intuition, toutes choses qui aident le sociologue dans sa démarche mais qui ne lui donnent pas l'assurance nécessaire pour conclure ? C'est ce que va nous apporter l'écart pondéré ou khi-deux :

	0 ou 1 redoublement	2 redoublements ou +	Total
--	------------------------	-------------------------	-------

Supérieure	2,8	9,5	12,3
Inférieure	3,1	10,6	13,7
Total	5,9	20,1	26,0

Le khi-deux total est de 26 : si l'on compare écarts bruts et écarts pondérés, on voit que pour toutes les cases la pondération s'est faite dans le sens de la réduction (puisque tous les écarts étaient inférieurs au théorique), mais que, même cette pondération faite, on se trouve encore avec une somme pondérée de 26 individus en écart à l'indépendance. Cet écart est-il suffisant pour chercher à interpréter le tableau, c'est à dire admettre qu'il y ait du sens au croisement, c'est à dire refuser de dire que nous sommes, comme dans le cas précédent, dans la situation d'indépendance ? La réponse sera guidée par l'existence de tables statistiques du khi-deux qui *a priori*, vont nous indiquer dans le cas présent d'un tableau 2×2 s'il est légitime ou non de penser trouver du sens au tableau, de trouver les écarts suffisamment forts pour qu'ils soient *significatifs*.

La table du khi-deux part de l'hypothèse que nous sommes dans une population du type de notre premier exemple où il n'y pas d'écart à l'indépendance : où il y a indépendance entre ce qui est en ligne et ce qui est en colonne. Cependant, du fait des fluctuations d'échantillonnage, on peut avoir la malchance de tirer un échantillon qui ne soit pas représentatif de cette population où l'hypothèse d'indépendance est vraie. Si l'on tire un assez grand nombre d'échantillons de la population, 100 par exemple, il arrivera malheureusement que pour un petit nombre d'entre eux, *bien que nous soyons en situation d'indépendance*, on constatera cependant des écarts à l'indépendance.

Ce que nous dit la table du khi-deux c'est que la somme des écarts pondérés atteindra au moins 0,5 dans la moitié des échantillons, ce qui

signifie qu'il y a habituellement quelques écarts dans les échantillons, même s'il y a indépendance et que :

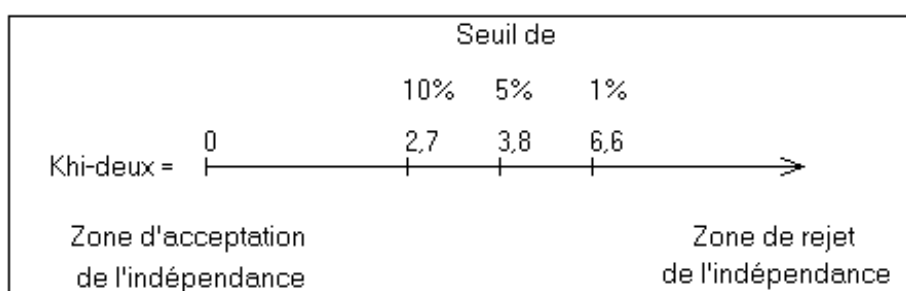
- dans 10 échantillons sur 100 le khi-deux atteindra la valeur de 2,7
- " 5 " 3,8
- " 1 " 6,6
- et même que dans 1 cas sur 1000, il atteindra 10,8.

Dans notre exemple, le khi-deux observé est de 26 : il n'est évidemment matériellement pas impossible que, avec une population de référence en situation d'indépendance, cette valeur soit obtenue, simplement du fait des fluctuations d'échantillonnage, peut-être dans un cas sur un million mais il devient plutôt raisonnable de prendre l'attitude humaine standard, que nous avons vue à l'œuvre plus haut (dans le cas de la panne de courant), qui est de refuser de croire au hasard trop exceptionnel et de trouver une autre explication qui sera ici tout simplement de rejeter l'hypothèse d'indépendance entre les deux questions. Ce serait vraiment un hasard trop incroyable d'avoir un khi-deux de 26 alors que dans un cas sur 100 seulement il atteint 6,6 par hasard. Il vaut mieux dire que s'il y a des écarts à l'indépendance, c'est que dans la population de référence *il y a un lien entre origine sociale et réussite scolaire* et que l'hypothèse d'indépendance n'est pas admissible.

Nous sommes maintenant autorisés à chercher à comprendre le sens de cette liaison qui nous signifie simplement que les étudiants d'IUT ont leur spécificité : s'ils ont choisi cette voie, c'est pour des raisons inverses selon leur origine sociale. Pour ceux d'origine sociale supérieure, s'ils n'ont pas entrepris d'aller dans des préparations aux grandes écoles, c'est que leur niveau scolaire ne le leur permettait pas. Des redoublements ont fait que les IUT sont pour eux un filet de rattrapage. Inversement pour les meilleurs éléments d'origine sociale inférieure, les IUT sont vus comme une filière moins aléatoire que l'université, plus proche culturellement de ce qui est connu dans leur milieu, facilement accessible grâce à leur bon niveau scolaire, moins

longue que les filières nobles et gage d'une bonne insertion professionnelle. Cette spécificité des IUT fait que le mauvais niveau scolaire est en attraction avec l'origine sociale supérieure et inversement pour l'origine sociale inférieure.

Résumons notre démarche par le diagramme KD1 où le khi-deux est croissant. S'il est nul ou très faible, à gauche du graphique, nous n'avons aucune raison de remettre en cause l'hypothèse d'indépendance, nous sommes en zone d'acceptation de l'hypothèse d'indépendance. Inversement à droite, si le khi-deux est très fort, il devient invraisemblable de croire à un hasard extraordinaire qui aurait conduit à cette valeur. Il est plus raisonnable de rejeter l'hypothèse d'indépendance et de chercher la signification de la liaison observée. Nous sommes en zone de rejet de l'indépendance. Entre les deux zones, la frontière est floue (nous allons étudier ce cas) et elle est balisée en général par les trois valeurs du hasard que nous avons cité : 10 chances sur 100 (soit une sur dix), 5 chances sur 100, 1 chance sur 100. C'est ce qu'on appelle les *seuils standards* de 10%, 5%, 1%.



KD1

IUT : sexe - réussite (nouveau codage)

Un exemple de khi-deux frontière nous est donné en reprenant le premier exemple IUT mais en isolant ceux qui sont strictement à

l'heure et ceux qui sont en retard, même d'un an. On a alors l'observation suivante :

	à l'heure	en retard	Total
Masculin	419	982	1401
Féminin	149	274	423
Total	568	1256	1824

Le PEM est de 5,9%, ce qui n'est pas très élevé : on peut se demander s'il y a une liaison significative bien que faible, ou si nous sommes dans le cas d'écarts aléatoires à l'indépendance ayant entraîné une liaison faible mais non significative. Pour trancher, nous utilisons le khi-deux qui est pour ce tableau de 4,3 soit, dans le graphique KD1, à la frontière entre les zones d'acceptation et de rejet de l'indépendance, entre les seuils de 5% et de 1%, ce qui veut dire que la probabilité d'avoir un khi-deux de 4,3 (s'il y a indépendance) n'arrivera par hasard que dans moins de 5% des cas (environ 3% des cas).

On peut tenir le raisonnement suivant : si j'accepte un risque de me tromper de l'ordre de 3% (et donc d'avoir raison dans 97% des cas, ce qui est assez satisfaisant), je peux dire que je suis légitimement en zone de rejet (de l'hypothèse d'indépendance) et dire que la liaison a du sens, que le tableau est significatif. Ici le sens est le suivant : il faut repérer que l'attraction se fait entre féminin et à l'heure, ce qui signifie que des filles strictement à l'heure (donc d'un bon niveau) sont plus souvent que les garçons poussés vers les IUT, vraisemblablement du fait que les familles ont moins d'ambitions pour les filles que pour les garçons. Les garçons de bon niveau sont moins tournés vers les IUT et davantage vers des études plus longues et plus prestigieuses. L'attraction est faible (PEM de 5,9%) mais cependant significative, comme l'on dit habituellement, *au seuil de 5%* ce qui signifie que le

khi-deux observé se trouve à droite de ce seuil. Ce seuil de 5% est la frontière retenue.

On peut étendre le raisonnement à d'autres valeurs : un khi-deux de 3 ou même de 2,7 sera réputé significatif au seuil de 10% et tout khi-deux supérieur à 6,6 sera réputé significatif au seuil de 1%. On voit bien que plus le khi-deux baisse, plus l'attraction baisse, plus on se rapproche de l'indépendance et il arrive un moment où il devient très risqué de rejeter l'hypothèse d'indépendance, d'absence de liaison entre lignes et colonnes. On déclarera alors, si le khi-deux est par exemple inférieur à 2,7 que le tableau est *non significatif* et l'on ne cherchera pas à rendre compte des faibles attractions observées qui sont aléatoires. Tant qu'on est à droite du seuil de 10%, le tableau est réputé significatif mais les risques pris sont plus ou moins grands, d'abord 10%, puis 5%, puis 1% et on ne cherche pas à donner des valeurs encore plus faibles : ces trois seuils sont dits *standards*.

4 Les tableaux qui ne sont pas 2×2

Dans un tableau 2×2 muni de ses marges, dès qu'une case est définie, les autres s'en déduisent par soustraction. Par exemple dans le tableau précédent muni de ses marges, si on ne donne que la case masculin - à l'heure :

	à l'heure	en retard	Total
Masculin	419		1401
Féminin			423
Total	568	1256	1824

les cas "masculin en retard" se déduisent de $1401 - 419 = 982$;

les cas "féminin à l'heure" par $568 - 419 = 149$;

les cas "féminin en retard" par $1256 - 982 = 274$.

On peut généraliser ce résultat : dans un tableau plus grand que 2×2 , il n'y a que la dernière ligne ou la dernière colonne qui puisse de déduire par soustraction des marges. Dans le cas d'un tableau 3×5 comme ci-dessous, il faut remplir les 8 cases marquées de l'intérieur du tableau pour que la 3e colonne et la 5e ligne puisse se déduire des marges : ce nombre de 8 est le nombre de colonnes = 3 diminué de 1 multiplié par le nombre de lignes = 5 diminué de 1 soit $2 \times 4 = 8$. C'est ce qu'on appelle communément le *degré de liberté d'un tableau*, c'est à dire le nombre de cases qu'il faut fixer pour que les autres se déduisent des marges.

	Colonne 1	Colonne 2	Colonne 3	Marge
Ligne 1	XXXXXXX	XXXXXXX		XXXXXXX
Ligne 2	XXXXXXX	XXXXXXX		XXXXXXX
Ligne 3	XXXXXXX	XXXXXXX		XXXXXXX
Ligne 4	XXXXXXX	XXXXXXX		XXXXXXX
Ligne 5				XXXXXXX
Marge	XXXXXXX	XXXXXXX	XXXXXXX	XXXXXXX

Plus le degré de liberté est grand est plus sont grandes les fluctuations du khi-deux dues à l'échantillonnage : les tables du khi-deux tiennent compte de cet aspect en augmentant les valeurs du khi-deux aux seuils standards en fonction du degré de liberté :

Table du khi-deux ⁴²			
Degré de liberté	Seuils		
	10%	5%	1%

⁴² on trouvera dans tout livre de statistique des tables plus étendues.

1	2,7	3,8	6,6
2	4,6	6,0	9,2
3	6,3	7,8	11,3
4	7,8	9,5	13,3
5	9,2	11,1	15,1
6	10,6	12,6	16,8
7	12,0	14,1	18,5
8	13,4	15,5	20,1
9	14,7	16,9	21,7
10	16,0	18,3	23,2
11	17,3	19,7	24,7
12	18,5	21,0	26,2
13	19,8	22,4	27,7
14	21,1	23,7	29,1
15	22,3	25,0	30,6
16	23,5	26,3	32,0
17	24,8	27,6	33,4
18	26,0	28,9	34,8
19	27,2	30,1	36,2
20	28,4	31,4	37,6

5 Exemple d'utilisation : PRAT89

Par exemple dans PRAT89 nous croisons l'âge du répondant avec le fait qu'il place l'émission "enquêtes du commissaire Maigret" dans ses émissions préférées. On a les résultats suivants en effectifs, en pourcentage en ligne, et en représentation graphique (graphique KD2) :

Age	Maigret : émission préférée		Total
	oui	non	

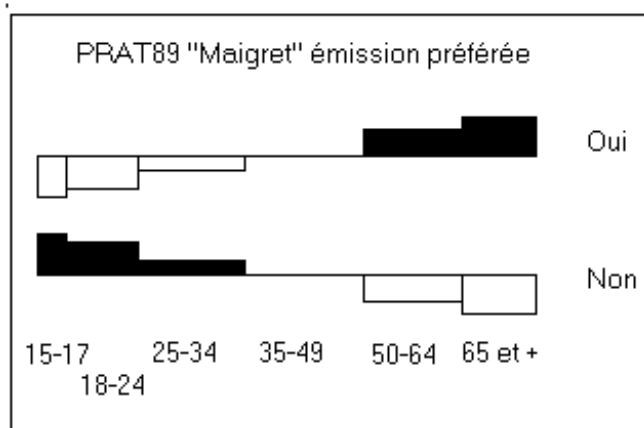
15-17	14	261	275
18-24	50	675	725
25-34	125	942	1067
35-49	190	1016	1206
50-64	219	754	973
65 et +	195	556	751

Total	793	4204	4997
-------	-----	------	------

Maigret : émission préférée

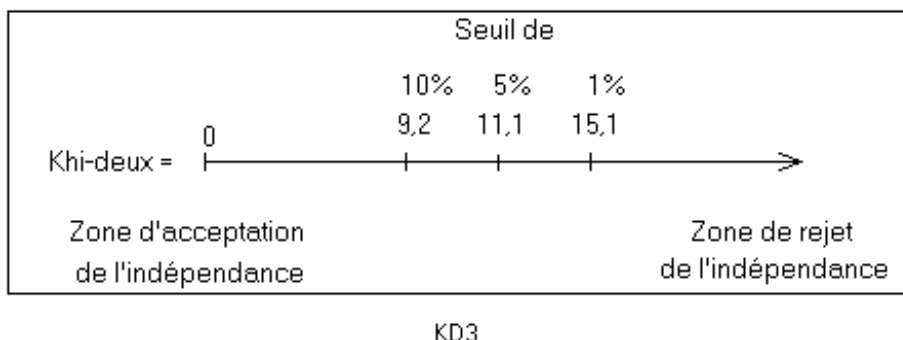
Age	oui	non	Total
15-17	5,1	94,9	100
18-24	6,9	93,1	100
25-34	11,7	88,3	100
35-49	15,8	84,2	100
50-64	22,5	77,5	100
65 et +	26,0	74,0	100

Total	15,9	84,1	100
-------	------	------	-----



KD2

Le PEM global du tableau est de 22,2% et le khi-deux de 170,9. Le degré de liberté est de 5 puisque c'est un tableau 2×6 . Dans le graphique d'acceptation/rejet, il faut remplacer les valeurs des seuils standards par ceux de la table de 5 degrés de liberté.



On voit immédiatement que la valeur de 170,9 se trouve tout à fait en zone de rejet de l'hypothèse d'indépendance : ce tableau est significatif (au seuil de 1%), il a du sens, on peut se servir des descriptions antérieures qui nous manifestent clairement que cette émission est préférée par les générations âgées. Plus l'âge augmente et plus la proportion de ceux qui placent l'émission dans leurs émissions préférées augmente.

Par contre si on croise cette émission avec le sexe de l'individu on a les résultats suivants :

Sexe	Maigret : émission préférée		Total
	oui	non	
Masculin	374	2030	2404
Féminin	419	2174	2593
Total	793	4204	4997
Sexe	Maigret : émission préférée		Total
	oui	non	

Masculin	15,6	84,4	100
Féminin	16,2	83,8	100
Total	15,9	84,1	100

La différence de pourcentage en faveur du sexe féminin est extrêmement faible (moins de 1%) : on est donc très près de l'indépendance. Le khi-deux du tableau est de 0,3 : en reprenant le schéma KD1 du khi-deux à 1 degrés de liberté, nous constatons que nous sommes en zone d'acceptation de l'indépendance. Les écarts sont trop faibles pour que nous remettons en cause cette hypothèse : il y a indépendance entre le sexe et le choix de cette émission.

6 Variations

Prenons un exemple de l'enquête FUM92:

	Fumeurs	Non fumeurs	Total
Mère fume	18	13	31
Mère ne fume pas	39	57	96
Total	57	70	127

Enquête pédagogique réalisée en 1992 (étudiants de licence).

On peut y repérer une attraction entre "mère fume" et "fumeurs" (PEM = 23,9) ; le khi-deux de ce tableau est de 2,9 soit, pour 1 degré de liberté, un tableau significatif au seuil de 10%. L'information est suffisante pour rejeter l'hypothèse d'indépendance. Le fait que la mère fume fait que l'étudiant est plus souvent fumeur.

Appliquons cependant la technique déjà utilisée du déplacement élémentaire : si un individu de la case en attraction (d'effectif 18) changeait de catégorie et que ce changement se répercute dans toutes

les cases de façon de laisser les marges inchangées, on aurait alors le déplacement élémentaire suivant que l'on ajoute au tableau précédent :

	Fumeurs	Non fumeurs
Mère fume	-1	+1
Mère ne fume pas	+1	-1

Le résultat du déplacement sera le suivant :

	Fumeurs	Non fumeurs	Total
Mère fume	17	14	31
Mère ne fume pas	40	56	96

Total	57	70	127

Le PEM a baissé puisqu'on a diminué la case en écart positif et qu'on s'est rapproché de l'indépendance, il est maintenant de 18,1% et le khi-deux de 1,6 : il n'est donc plus significatif. Il a donc suffi d'un seul déplacement pour qu'on ne puisse plus exclure l'indépendance.

Continuons ce travail dans les deux sens, tout d'abord en se rapprochant le plus possible de l'indépendance qui est de 13,9 pour la première case du tableau ("Mère fume - fumeurs"). Comme nous sommes dans un tableau à un seul degré de liberté, une fois cette case fixée, toutes les autres sont définies. L'entier le plus proche est 14. Voici le tableau le plus proche de l'indépendance et ses voisins jusqu'au tableau le plus extrême par déplacement élémentaire avec le PEM et le Khi-deux :

Tableau	PEM	Khi-deux
14 17 43 53	0,5%	0,001
tableau le plus proche de l'indépendance		
15 16 42 54	6,4%	0,204

16 15 41 55	12,2%	0,751	
17 14 40 56	18,1%	1,644	
	-----	2,7 : seuil de 10%	
18 13 39 57	23,9%	2,881	tableau observé
	-----	3,8 : seuil de 5%	
19 12 38 58	29,8%	4,463	
20 11 37 59	35,6%	6,391	
	-----	6,6 : seuil de 1%	
21 10 36 60	41,5%	8,663	
-----	4 tableaux intercalaires sautés de première case 22, 23, 24 et 25		
26 5 31 65	70,7	25,201	
-----	4 tableaux intercalaires sautés de première case 27, 28, 29 et 30		
31 0 26 70	100%	50,364	tableau extrême (loin de l'indépendance)

Examinons ce panorama de tableaux : le tableau "14" (on peut désigner les tableaux par leur première case) est le plus proche de l'indépendance et le 31 le plus éloigné. Ceci se repère par un PEM qui est très proche de 0 pour le premier et égal à 100% pour le dernier. Entre ces deux extrêmes se situe l'observé : le 18. Notons bien qu'un déplacement vers l'indépendance (vers le 17) rend le tableau non-significatif, qu'un déplacement vers la liaison maximum (jusqu'au 19) le rend significatif au seuil de 5% et que deux déplacements (jusqu'au 21) le rendent significatif au seuil de 1%. D'un déplacement à l'autre,

le PEM croit d'environ 6% à chaque fois, le khi-deux d'une manière moins régulière (d'un peu plus d'une unité autour de l'observation).

Ces fortes variations sont un effet du faible effectif. Un déplacement de 1 individu fait franchir les frontières de significativité du khi-deux et fait varier beaucoup le PEM. Si nous avons, avec la même structure, des effectifs plus forts, les variations seraient beaucoup plus faibles. Prenons par exemple le tableau 16, non significatif (khi-deux de 0,751 et PEM=12,2%). Pour conserver la même structure, multiplions simplement tous les effectifs par 10 et examinons ses voisins pour un déplacement élémentaire :

Tableau	PEM	Khi-deux
159 151 411 549	11,6%	6,808
160 150 410 550	12,2%	7,511
tableau de référence (effectifs multipliés par 10)		
161 149 409 551	12,8%	8,248

Le PEM maintenant est peu modifié, il reste aux alentours de 12%, et le khi-deux qui varie lui aussi peu, d'une unité environ, a été multiplié par 10 pour le tableau de référence : d'un khi-deux de 0,75 on est passé à un khi-deux de 7,5. Il n'y a là aucun paradoxe⁴³ : une même liaison sur un petit effectif est réputée non significative ; avec un effectif 10 fois plus grand, l'information est plus grande, le khi-deux est plus grand et devient suffisant ici pour que le tableau soit

⁴³ Si l'écart est multiplié par 10, le rapport écart/théorique reste inchangé puisque le numérateur et le dénominateur de la fraction ont été multipliés par 10.

réputé significatif au seuil de 1%. Ce n'est pas la force de la liaison qui a changé : le PEM est indifférent aux effectifs, c'est l'information qui est devenue suffisante pour entraîner une autre décision. Cela n'est d'ailleurs pas étranger au sens commun : si l'on observe qu'une personne observée sur 10 est rousse, il est difficile d'extrapoler ce résultat. Si l'on en a observé 10 sur une population de 100, la prédiction devient meilleure, et si on arrive à 100 sur 1000 la même prédiction devient excellente. Un faible écart à l'indépendance peut facilement être assimilé à des fluctuations d'échantillonnage quand l'effectif observé est faible. Il peut être mesuré avec précision bien que faible, quand l'effectif est important. Le khi-deux n'est que le reflet de l'effectif car on peut le considérer comme un effectif pondéré.

7 Règles de bonne utilisation

On trouve dans la plupart des livres de statistiques des règles de précaution qui entourent l'utilisation du test du Khi-deux. Pour voir où réside une éventuelle difficulté, étudions l'exemple suivant, tableau croisé issu de l'enquête REL86 et qui croise le sexe avec la question "Quand vous étiez enfant, vos parents étaient-ils pratiquants ?" avec les modalités de réponse suivantes (et leur effectif) : votre père et votre mère (653) ; votre père seulement (20) ; votre mère seulement (262) ; ni votre père ni votre mère (560) ; était orphelin (9) ; non-réponse (26). Il y a par ailleurs dans l'enquête 726 hommes, 803 femmes (et une non-réponse à cette question). Le tableau est le suivant :

Sexe	Père +Mère	Père	Mère	Ni Père ni Mère	Orph	Non-rép.	Total
Masculin	309	8	128	265	5	11	726
Féminin	344	12	134	295	4	14	803
Non-réponse	0	0	0	0	0	1	1

-

Total	653	20	262	560	9	26	1530
-------	-----	----	-----	-----	---	----	------

Le khi-deux de ce tableau est de 58,9, ce qui, avec un degré de liberté de 10, nous met en zone de rejet de l'hypothèse d'indépendance à tous les seuils standards (le seuil de 1% correspond à une valeur de 23,2), ce qui indiquerait donc qu'il y aurait une relation entre le sexe et la croyance des parents. Comme ceci semble absurde puisqu'il est peu probable que le sexe de l'enfant dépende de la pratique religieuse des parents, examinons les pourcentages en ligne.

Sexe	Père +Mère	Père	Mère	Ni Père ni Mère	Orph	Non-rép.	Total
Masculin	42,6	1,1	17,6	36,5	0,7	1,5	100
Féminin	42,8	1,5	16,7	36,7	0,5	1,7	100
Non-réponse	0	0	0	0	0	100	100
Total	42,7	1,3	17,1	36,6	0,6	1,7	100

On constate que les deux premières lignes de pourcentage sont pratiquement identiques entre elles et avec le total : il n'y a pratiquement pas d'écarts au pourcentage moyen, donc d'écarts à l'indépendance, donc il ne devrait pas y avoir d'écarts pondérés et les contributions au khi-deux devraient être insignifiantes. Nous examinons ce point en regardant le tableau des contributions au khi-deux de chaque case.

Sexe	P+M	P	MNi	P.ni	M.	Orph	NR.
Masculin	0,0	0,2	0,1	0,0	0,1	0,1	
Féminin	0,0	0,2	0,1	0,0	0,1	0,0	
Non-réponse	0,4	0,0	0,2	0,4	0,0	56,9	

On voit que les contributions sont effectivement insignifiantes sauf sur la dernière case qui croise les non-réponses aux deux questions et

qui correspond à un seul individu. Reprenons le calcul du khi-deux pour cette case :

Observé : 1

Théorique : $26 \times 1 / 1530 = 0,01699$

Ecart à l'indépendance = $1 - 0,01699 = 0,98300$

Ecart au carré = 0,96630

Khi-deux = $0,96630 / 0,01699 = 56,9$

On voit que cet écart inférieur à 1 individu entraîne un fort khi-deux du fait du dénominateur de la division qui est très faible : diviser par 0,01 revient à multiplier par 100. L'amplification liée à la pondération par l'effectif théorique devient une majoration excessive, exagérée. Pratiquement tout le khi-deux du tableau (96,6%) vient de cette case.

On constate sur cet exemple que, lorsque l'on divise par des effectifs théoriques très faibles, les contributions au khi-deux résultantes sont exagérées et risquent d'induire en erreur. Ceci nous pousse donc à considérer le test du khi-deux comme un garde-fou, un clignotant qui s'allume dans certaines circonstances particulières et qui doit mettre en éveil l'utilisateur. Si on constate que le clignotant s'allume à tort, on l'éteint. Concrètement, puisque ce risque vient de trop faibles effectifs théoriques, il faut se méfier de ceux-ci quand ils sont inférieurs à 1 ce qui arrive quand l'une des marges est faible⁴⁴. La solution est simple, il suffit de procéder à une opération de recodage : ici il faut agréger la non-réponse à l'une des deux modalités ou l'éliminer. On recode ici avec le sexe masculin⁴⁵. On a alors le tableau suivant, très peu différent du précédent.

⁴⁴ On retrouve ainsi la précaution la plus indiquée dans les manuels qui est de se méfier des tableaux où les théoriques sont inférieurs à 1.

⁴⁵ L'expérience montrant que ce cas d'oubli est le plus fréquent.

Sexe	Père +Mère	Père	Mère	Ni Père ni Mère	Orph	Non-rép.	Total
Masculin	309	8	128	265	5	12	727
Féminin	344	12	134	295	4	14	803
Total	653	20	262	560	9	26	1530

Le khi-deux de ce tableau est de 0,9 : il y a indépendance entre les deux questions car le degré de liberté est maintenant de 5 et au seuil de 10% on a une valeur pour le khi-deux de 9,24. Le khi-deux observé est en zone d'acceptation de l'indépendance.

En conclusion, et plutôt que de donner des règles de cuisine qui s'apparentent rapidement à des règles de magie, nous invitons l'utilisateur à être prudent quand une lecture des pourcentages ou du graphique des écarts semble en opposition avec le test du khi-deux. Quand par exemple les profils des réponses sont plats alors que le khi-deux semble indiquer un rejet de l'indépendance. Dans cette situation, il faut regarder les contributions au khi-deux et voir si l'une d'entre elles ne serait pas liée à un faible effectif théorique (inférieur à 1). Si c'est le cas, il faut recoder les faibles effectifs en les agrégeant à la modalité la plus proche par le sens.

Le khi-deux est un indicateur de l'information apportée par chaque case, c'est un outil au service de l'interprétation : ce n'est pas une guillotine qui s'abat sur un tableau pour l'éliminer. Il faut observer le contenu du tableau et comprendre d'où viennent des incohérences éventuelles.

La procédure de recodage est d'ailleurs un outil qui sert non seulement pour éviter les aberrations comme celles vues à l'instant, mais également pour rendre cohérent un khi-deux et une interprétation dans l'autre sens, c'est à dire quand l'utilisateur juge que le tableau est le reflet d'une liaison qui a du sens entre questions alors que le test n'est pas significatif, c'est à dire qu'il nous suggère une acceptation de

l'hypothèse d'indépendance. En effet, si le khi-deux ne nous invite pas à rejeter l'hypothèse d'indépendance, on a vu qu'avec un effectif plus fort et le même niveau de liaison (PEM), le khi-deux deviendrait significatif. On peut atteindre ce résultat en réduisant le degré de liberté par diminution du nombre de lignes et de colonnes du tableau (par recodage). Evidemment, si après ce recodage, le khi-deux reste non-significatif, il faut abandonner cette hypothèse si l'on n'a pas de nouvelles observations. On va voir cette procédure à l'œuvre dans l'exemple suivant.

8 Exemple d'utilisation : FUM97.

Soit donc l'enquête FUM97 dont nous avons parlé en introduction : on interroge les 225 étudiants d'un deug de sociologie et on leur pose trois questions : s'ils fument ou non (au sens de l'INSEE, c'est à dire au moins une cigarette par jour), leur sexe et l'origine sociale de leurs parents.

Soit le premier croisement entre le sexe et le fait de fumer :

Croisement question SEX et question FUM

Le Khi-deux du tableau est de **8.5 ***⁴⁶ PEM=34%**

COL:	FUM	FUM		FUM	FUM			FUM	FUM	
Modal.	oui	non	TOT	oui	non	TOT		oui	non	
MASC	29	17	46	63.0	37.0	100	20.4	MASC	+	-
FEMI	70	109	179	39.1	60.9	100	79.6	FEMI	-	+
TOT	99	126	225	44.0	56.0	100	100			

Il y a donc un lien entre le fait de fumer et le sexe : le sexe masculin fume plus que le sexe féminin (63% contre 39%). Etudions le lien avec la catégorie socioprofessionnelle :

⁴⁶ On note souvent le résultat du test du khi-deux de la manière symbolique suivante : *** significatif au seuil de 1% ; **=5% ; *=10% ; ns=non-significatif.

Croisement question CSP et question FUM

Le Khi-deux du tableau est de 8.1 ns PEM=14%

COL:	FUM	FUM		FUM	FUM			FUM	FUM
Modal.	oui	non	TOT	oui	non	TOT		oui	non
CSP 0	0	4	4	0.0	100.0	100	1.8	CSP0	- + NR
CSP 1	2	3	5	40.0	60.0	100	2.2	CSP1	- + Agri
CSP 2	11	8	19	57.9	42.1	100	8.4	CSP2	+ - ArtCo
CSP 3	25	23	48	52.1	47.9	100	21.3	CSP3	+ - Csup
CSP 4	11	12	23	47.8	52.2	100	10.2	CSP4	+ - Inter
CSP 5	14	28	42	33.3	66.7	100	18.7	CSP5	- + Empl
CSP 6	36	48	84	42.9	57.1	100	37.3	CSP6	- + Ouvr
TOT	99	126	225	44.0	56.0	100	100		

Si l'on examine les écarts en lecture rapide, on voit que trois CSP fument plus que les autres, les Artisans-commerçants, les cadres et professions intellectuelles supérieures et les professions intermédiaires. Les autres, en particulier employés et ouvriers qui représentent 60% du total, fument moins que la moyenne. Comme le tableau n'est pas significatif et que l'on voit une logique d'opposition qui a du sens, on tente un regroupement en deux catégories de CSP : une CSP sup. (ArtCom, Csup, Inter) et une CSP inf. (empl, ouvr et le reste). On a alors le résultat suivant :

Croisement question CSP et question FUM

Le Khi-deux du tableau est de 4.1 ** PEM=15%

COL:	FUM	FUM		FUM	FUM			FUM	FUM
Modal.	oui	non	TOT	oui	non	TOT		oui	non
CSPsup	47	43	90	52.2	47.8	100	40.0	CSPsup	+ -
CSPinf	52	83	135	38.5	61.5	100	60.0	CSPinf	- +
TOT	99	126	225	44.0	56.0	100	100		

Cette fois, le résultat est significatif, ce qui montre l'efficacité du regroupement par réduction du degré de liberté. Ceci signifie qu'avec un codage moins fin, les effectifs sont suffisants pour apporter une information fiable. Il y a une liaison entre CSP et fait de fumer, la classe supérieure fumant davantage.

On constate que d'un recodage à l'autre, le PEM est resté stable. Comme indice de liaison, le PEM est indépendant de l'effectif et peu sensible aux recodages.

On va pouvoir maintenant prendre en compte les trois questions en même temps en regardant la liaison entre le sexe et le fait de fumer pour deux sous-populations (analyse dite *multivariée*). Dans un premier tableau on étudiera la sous-population de classe supérieure (90 individus) puis dans un deuxième, celle de classe inférieure (135 individus). Soit la population de classe supérieure :

Variable test CSP modalité CSPsup

Croisement question SEX et question FUM

Le Khi-deux du tableau est de 0.1 ns PEM=6%

COL:	FUM	FUM		FUM	FUM			FUM	FUM
Modal.	oui	non	TOT	oui	non	TOT		oui	non
MASC	11	9	20	55.0	45.0	100	22.2	MASC +	-
FEMI	36	34	70	51.4	48.6	100	77.8	FEMI -	+
TOT	47	43	90	52.2	47.8	100	100		

Garçons et filles de classe supérieure ont à peu près le même comportement vis-à-vis du tabac, ils fument plus que la classe inférieure (52% contre 39%) mais les deux sexes accusent peu de différence en termes d'écart au pourcentage moyen. Le khi-deux n'est pas significatif, le PEM faible : on peut admettre l'hypothèse d'indépendance.

Du point de vue de l'interprétation sociologique, ceci est cohérent avec une tendance à l'unification des comportements des sexes qui vont en se rapprochant, particulièrement dans la classe supérieure. Garçons et filles s'habillent de la même façon, ont les mêmes loisirs, souvent les mêmes désirs en termes professionnels et cette tendance est plutôt encouragée par leur milieu. Il est donc normal qu'ils aient le même comportement vis à vis de la consommation de tabac. Il n'en est pas de même dans la classe inférieure :

Variable test CSP modalité CSPinf

Croisement question SEX et question FUM

Le Khi-deux du tableau est de 12.8 *** PEM=50%

COL:	FUM	FUM		FUM	FUM			FUM	FUM	
Modal.	oui	non	TOT	oui	non	TOT		oui	non	
MASC	18	8	26	69.2	30.8	100	19.3	MASC	+	-
FEMI	34	75	109	31.2	68.8	100	80.7	FEMI	-	+
TOT	52	83	135	38.5	61.5	100	100			

On voit une différence très nette entre le comportement des garçons qui fument à 69% (c'est à dire plus que garçons et filles de la classe supérieure) et les filles qui ne fument qu'à 31%. Là, l'uniformisation des comportements n'est pas pratiquée, en matière éducative en particulier : ce qui est admis pour les garçons (le tabac en particulier mais bien d'autres choses) ne l'est pas pour les filles. La différence est significative, la liaison est forte.

Conclusion

On notera sur cet exemple que la lecture rapide est évidemment trop rapide dans les deux derniers tableaux et que d'autre part un tableau réputé non-significatif par le khi-deux est tout à fait interprétable. Qu'il n'y ait pas de différence entre garçons et filles en classe supérieure a du sens (le terme "non-significatif" signifie simplement que l'écart à l'indépendance n'est pas suffisant pour rejeter l'hypothèse d'indépendance). Ce qui doit guider l'interprétation, c'est la signification sociologique donnée au croisement. On observe les données pour y trouver du sens, et on se sert à cette fin des diverses méthodes présentées au chapitre premier : ensuite le test du khi-deux n'est qu'un garde fou pour éviter des accidents, un clignotant qui attire notre attention sur de faibles effectifs.

Faire de la sociologie aujourd'hui, c'est étudier des données empiriques : données d'archives, entretiens, textes, observations, statistiques, et données d'enquêtes. Ces dernières nécessitent des techniques particulières pour être décrites. On a présenté ici des techniques simples mais qui, du fait de leur informatisation, peuvent traiter des enquêtes importantes. Comme on l'a dit en introduction, il

existe des techniques plus complexes qui nécessitent des développements particuliers (analyse factorielle et régression).

Bibliographie

- BERTIN, Jacques, 1977, *La graphique et le traitement graphique de l'information*, Paris, Flammarion, 1977. Ce livre est à consulter par toute personne voulant réfléchir sur des représentations graphiques.
- NOVI, Michel, 1988, *Pourcentages et tableaux statistiques*, Paris, Presses Universitaires de France, coll. "Que sais-je ?", n°3337. Le but de ce livre est de faire le point sur des connaissances et des pratiques souvent jugées trop élémentaires pour être enseignées.
- ROUANET Henry, LEROUX Brigitte, BERT Marie-Claude, 1987, *Statistique en sciences humaines : procédures naturelles*, Paris, Dunod, et :
- ROUANET Henry, BERNARD Jean-Marc, LEROUX Brigitte, 1990, *Analyse inductive des données*, Paris, Dunod. Ces deux livres constituent des références qui, bien que d'accès difficile par rapport au niveau du présent livre, devront être utilisés en particulier par les enseignants et les chercheurs qui souhaiteraient approfondir ces questions.
- SINGLY, François de, 1992, *L'enquête et ses méthodes ; le questionnaire*, Paris, Nathan, coll. 128. Savoir lire et produire des chiffres est considéré comme une des compétences à posséder pour les étudiants en sciences sociales. Ce livre veut être un soutien à cet apprentissage de l'enquête par questionnaire.

Enquêtes utilisées

- FUM92 Enquête "pédagogique" sur 127 étudiants de licence de sociologie en 1992
- FUM97 Enquête "pédagogique" sur 225 étudiants d'un deug de sociologie en 1997
- PRAT89 *Nouvelle enquête sur les pratiques culturelles des français en 1989*, Paris, 1990, La Documentation française. Enquête réalisée par le Ministère de la culture et portant sur 4997 individus.
- LATIN94 Enquête faite en 1994 dans le cadre de l'enseignement de méthodologie du Deug de Paris V en collaboration avec François de Singly. Elle est centrée sur le choix du latin. 1058 répondants.
- PANEL89 Cohorte d'élèves suivis par les services statistiques du Ministère de l'éducation nationale depuis leur entrée en 6^e en 1989. Ici, sous-population de 1865 individus.
- REL86 Guy Michelat, Julien Potel, Jacques Sutter, Jacques Maître, *Les Français sont-ils encore catholiques ?*, Paris, Cerf, 1991 Enquête portant sur 1530 répondants dont a été tiré ce livre.

Logiciel

TRIDEUX qui a été utilisé pour les analyses de ce livre est un logiciel libre réalisé par l'auteur. Il est cédé gratuitement et sa reproduction est libre. Informations auprès de l'auteur à l'adresse : cibois@francenet.fr